AD_____

Award Number: W81XWH-09-1-0062

TITLE: Image Based Biomarker of Breast Cancer Risk: Analysis of Risk Disparity Among Minority Populations

PRINCIPAL INVESTIGATOR: Fengshan Liu

CONTRACTING ORGANIZATION: Delaware State University
                            Dover, DE 19901

REPORT DATE: February 2015

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
               Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
                        Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
Feb 2015	Final	1Mar2009 - 30Nov2014

4. TITLE AND SUBTITLE

Image Based Biomarker of Breast Cancer Risk:

Analysis of Risk Disparity among Minority Populations

5a. CONTRACT NUMBER
W81XWH-09-1-0062

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Fengshan Liu, Xiquan Shi, Charlie, Wilson, Dragoljub Pokrajac, Predrag Bakic,

Andrew Maidment

email: fliu@desu.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Delaware State University
1200 N DuPont Hwy
Dover, DE 19901

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, MD 21702-5014

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This DoD BCRP HBCU Partnership Training project was designed to enhance Delaware State University (DSU) breast cancer research resources. We have performed a multifaceted training of a cadre of DSU faculty to study breast cancer and establish an independent research program at DSU through a joint DSU – University of Pennsylvania (UPENN) research project focused on breast cancer risk disparity in minority populations. This introduction gives a brief overview of the training and research activities performed on this project, which are detailed in the body of the report.

15. SUBJECT TERMS
Breast Cancer, Risk Disparity, Minority Population, Image-Based Biomarker, Training Program

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	UU	105	**19b. TELEPHONE NUMBER** (include area code)

Table of Contents

1. Introduction

This DoD BCRP HBCU Partnership Training project was designed to enhance Delaware State University (DSU) breast cancer research resources. We have performed a multifaceted training of a cadre of DSU faculty to study breast cancer and establish an independent research program at DSU through a joint DSU – University of Pennsylvania (UPENN) research project focused on breast cancer risk disparity in minority populations. This introduction gives a brief overview of the training and research activities performed on this project, which are detailed in the body of the report.

During Y1 and Y2, we focused on training and research. In Y1, DSU faculty working on this project took various classes at UPENN, focused at Breast Cancer Epidemiology, Biology, and Imaging (detailed in Section 2.1.1). In addition to these didactic classes, we organized a DSU-UPENN Breast Cancer Basics Seminar Series, held bi-weekly at DSU (Section 2.1.2). Regarding the research project, we have reviewed the database of anonymized clinical multimodality breast images (mammography, tomosynthesis, MRI, ultrasound, and PET), previously acquired within the NIH Program Project Grant supported clinical study at UPENN. Goal of the review was to familiarize with standard DICOM format of medical images, and to develop a Matlab code for extracting and processing metadata from images (Section 2.2). We have also we submitted applications and obtained the UPENN IRB and ACRIN approvals for the proposed transfer of mammography images from all minority women and their Caucasian controls from the ACRIN DMIST trial, to be used in our cancer risk disparity study (Section 2.2.1).

In Y3, we continued our partnership on breast cancer training and research. The bi-weekly DSU-UPENN Breast Cancer Basics Seminar Series has been continued at DSU. In addition, DSU faculty attended a number of breast cancer related seminars at UPENN. We have also continued with the transfer and analysis of requested images from the ACRIN DMIST database. Specifically, we integrated the ACRIN data into a clinical MIRC database and started with the analysis of image based biomarkers of breast cancer risk (Section 2.2.1). In addition, we worked on the refinement of the UPENN breast anatomy simulation method (Section 2.2.4). Our new design of the software breast phantom and its GPU implementation, as well as the proposed partial volume representation, have been published in several conference and journal papers (listed in Chapter 4).

In Y4 we finalized the ACRIN DMIST data transfer and resolved issues with various batch transfers. We performed a preliminary query of the ACRIN data aimed at identifying the prevalence of women with incomplete visualization of the breast (Section 2.2.1). A novel breast image registration method has been proposed to obtain a composite mammogram from several images with partial breast coverage, for the purpose of accurate breast density estimation (Section 2.2.3; publication listed in Chapter 4). We developed a code to estimate the breast cancer risks using the demographic metadata from the ACRIN cases (Section 2.2.1). We estimated mammographic breast density from ACRIN DMIST images using the software developed at the UPENN (Section 2.2.4). We have continued with the refinement of the UPENN software breast phantom (Section 2.2.4, publications listed in Chapter 4). We have also designed a method to improve thickness control of the Cooper's ligaments in the simulation algorithm by reducing "dents" on the ligaments' surface (Section 2.2.4). We have submitted the first proposal from this project, NIH R01 on the continued development of the breast anatomy and imaging simulation (PIs: Bakic and Pokrajac). The proposal was scored at 41% but not funded.

During the no-cost extension year, we have improved the use of the project database by constructing a web-based data center (Section 2.2.1). We have also completed the analysis of the ACRIN DMIST images. By merging source data, converted data and computed data together to create a relational database and developing facilitating functions, we implemented various kinds of data selection requirements in terms of SQL statements, and apply further statistics routines to discover more hidden correlations among different quantities. Our manuscript about cancer risk related results are being prepared for publication. We also updated the breast imaging simulation pipeline and a computer demo of real-time simulation, and developed proof for

computational complexity of the simulation algorithm and demonstrated its asymptotic efficiency (section 2.2.4; publications listed in Chapter 4). Finally, our proposal to NIH INBRE program, on the analysis and simulation of breast small scale tissue structures (PI: Pokrajac) was funded in February 2014. The following sections (entitled based upon the original objectives from the grant Statement of Work) detail the performed work and its deliverables.

2. Body

(*The following introductory paragraphs and the grant objectives – used as headings in the following text – were copied from the original grant proposal.*)

With this funded project, we will enhance DSU breast cancer research resources by: improving our expertise in translational and clinical breast cancer research; developing methods for computing image-based biomarkers for breast cancer risk, as well as methods for biomarker analysis of risk disparity; developing a database of clinical biomarkers computed from images of minority women; refining the existing and developing novel data mining techniques to determine the relationship between risk and image-based biomarkers. The improvement will support further growth of a sustained breast cancer research program at DSU and help establish us as a mid-Atlantic center for analysis of breast cancer risk and risk disparity among minority women.

The specific objectives of this training program include: (1) extending the skills of a select cadre of DSU faculty, so that they may become accomplished, influential and competitive breast cancer researchers; (2) establishing an independent breast cancer research program at DSU by performing a joint DSU–UPENN research project focused on breast cancer risk disparity in minority populations; and (3) producing a corpus of high quality published work and develop a portfolio of independently funded research grants at DSU to support a sustained breast cancer research program.

2.1 Objective 1

Objective 1 (from SOW): Extend the skills of a select cadre of Delaware State University (DSU) faculty, so that we may become accomplished, influential, and competitive breast cancer researchers.

2.1.1 Specific Training for DSU Faculty
Complementing Individual Scientific Backgrounds

Fall 2009 Semester

Fengshan Liu and Xiquan Shi took:

BE 483-401 2009C. Molecular Imaging
Course content includes: Structure of an atom, electromagnetic radiation, electron orbitals, the nucleus; radioactive decay, interactions of radiation with matter, X-ray imaging instrumentation; interactions of x-rays with tissue, computed tomography, X-ray contrast media, ultrasound image, and magnetic resonance imaging (MRI).

Dragoljub Pokrajac and Charlie Wilson took:

EP 801 Fundamentals of Epidemiologic Study Designs

This course is a series of lectures designed to teach basic principles of epidemiologic research design. Lectures include the following topics: definitions of epidemiology; measures of disease frequency; measures of effect

and association; epidemiologic study designs, both experimental and non-experimental; and an overview of analysis of epidemiologic studies.

Spring 2010 Semester

Xiquan Shi took:

CAMB 512-001 2010A. Cancer Biology & Genetic: Cancer Biology and Genetics
The course objective is to introduce the students to important and timely concepts in Cancer Biology and Cancer Genetics. The lectures are organized into four broad thematic groups: A) Cell-Autonomous Mechanisms (e.g., tumor suppressor and oncogene function, DNA repair pathways, senescence, apoptosis); B) Non Cell-Autonomous Mechanisms (e.g., tumor microenvironment, hypoxia, angiogenesis); C) Organ Systems (e.g., pancreatic cancer, hematopoietic malignancies); and D) Therapeutic Approaches (e.g. protein kinase inhibitors, immunotherapy, radiation therapy). The organizers, along with faculty from the School of Medicine, the Wistar Institute and CHOP, with expertise in the corresponding areas provide lectures for the course. The students are expected to present, and participate in discussions of one or more key recent papers at Journal Clubs that are held at the end of each thematic group. There will be mid-term and final exams of short essays relevant to the lectures.

Fengshan Liu and Dragljub Pokrajac took:

BE545/CIS 537   Biomedical Image Analysis
This course covers the fundamentals of advanced quantitative image analysis that apply to all of the major and emerging modalities in biological/biomaterials imaging and in vivo biomedical imaging. While traditional image processing techniques will be discussed to provide context, the emphasis will be on cutting edge aspects of all areas of image analysis (including registration, segmentation, and high-dimensional statistical analysis). Significant coverage of state-of-the-art biomedical research and clinical applications will be incorporated to reinforce the theoretical basis of the analysis methods.

Graduate Courses taken at UPENN

Spring 2011 Course

Dr. Charlie D. Wilson took:

GCB 535 Intro to Bioinformatics

Course Description: The course provides a broad overview of bioinformatics and computational biology as applied to biomedical research. Course material will be geared towards answering specific biological questions ranging from detailed analysis of a single gene through whole-genome analysis, transcriptional profiling, and systems biology. The relevant principles underlying these methods will be addressed at a level appropriate for biologists without a background in computational sciences. This course should enable students to integrate modern bioinformatics tools into their research program.

Spring 2010 Semester

Xiquan Shi took:

CAMB 512-001 2010A. Cancer Biology & Genetic: Cancer Biology and Genetics
The course objective is to introduce the students to important and timely concepts in Cancer Biology and Cancer Genetics. The lectures are organized into four broad thematic groups: A) Cell-Autonomous Mechanisms (e.g., tumor suppressor and oncogene function, DNA repair pathways, senescence, apoptosis); B) Non Cell-Autonomous Mechanisms (e.g., tumor microenvironment, hypoxia, angiogenesis); C) Organ Systems
(e.g., pancreatic cancer, hematopoietic malignancies); and D) Therapeutic Approaches (e.g. protein kinase inhibitors, immunotherapy, radiation therapy). The organizers, along with faculty from the School of Medicine, the Wistar Institute and CHOP, with expertise in the corresponding areas provide lectures for the course. The students are expected to present, and participate in discussions of one or more key recent papers at Journal Clubs that are held at the end of each thematic group. There will be mid-term and final exams of short essays relevant to the lectures.

Spring 2010 Semester

Fengshan Liu and Dragoljub Pokrajac took:

BE545/CIS 537 Biomedical Image Analysis
This course covers the fundamentals of advanced quantitative image analysis that apply to all of the major and emerging modalities in biological/biomaterials imaging and in vivo biomedical imaging. While traditional image processing techniques will be discussed to provide context, the emphasis will be on cutting edge aspects of all areas of image analysis (including registration, segmentation, and high-dimensional statistical analysis). Significant coverage of state-of-the-art biomedical research and clinical applications will be incorporated to reinforce the theoretical basis of the analysis methods.

Spring 2010 Semester

Charlie Wilson took:

GCB/CAMB 752 SEMINAR IN GENOMICS
Recent papers from the primary genomics literature will form the core material for the course. Each 3-hr session will feature a major topic or set of related topics in Genomics, with student presentations (usually two per session) centered on papers selected within the topic area(s). While the "presenting" student will give a 10- 15 min introduction to the paper and will show PowerPoint slides of the data in the paper, all students in the class are expected to have read and to be prepared to discuss the papers presented

Spring 2011 Course

Charlie Wilson took:

GCB 535 Intro to Bioinformatics

Course Description: The course provides a broad overview of bioinformatics and computational biology as applied to biomedical research. Course material will be geared towards answering specific biological questions ranging from detailed analysis of a single gene through whole-genome analysis, transcriptional profiling, and systems biology. The relevant principles underlying these methods will be addressed at a level appropriate for biologists without a background in computational sciences. This course should enable students to integrate modern bioinformatics tools into their research program

o   **Augment the faculty training by frequent communications with collaborating mentors and other renowned breast cancer researchers, by: (Y1-4)**

2.1.2 Biweekly DSU-UPENN Breast Cancer Seminar Series

DSUPENN Breast Cancer Seminar Series

DSUPENN Breast Cancer Seminar Series are organized to provide training in breast cancer research to DSU faculty including Fengshan Liu, Xiquan Shi, Charlie Wilson and Dragoljub Pokrajac and students at Delaware State University. Invited speakers of the biweekly seminar series include nationally renowned breast cancer researchers from UPENN Medical School, nearby hospitals and other institutions.

Speakers:
Invited speakers of the biweekly seminar series include nationally renowned breast cancer researchers from UPENN Medical School, nearby hospitals and other institutions.

Before October 9, 2009: We named the seminar series as Biweekly DSU-UPENN Breast Cancer Basics Seminar Series. Contents including: Breast Cancer Risk Factors; The Biology of Breast Cancer (I, II and III); and Cancer Imaging.

Date: July 10, 2009
Seminar title: Inverse Cell Biology Overview
Speaker: Charlie Wilson, DSU
Description: This presentation provided an introduction to the general components of animal cells and the specific types of cells found in breast tissue. The basics of cell growth dynamics and changes associated with cancer were also addressed. The role of gene products from proto-oncogenes and tumor suppressor genes in cancer were discussed.

Date: July 13, 2009
Seminar title: Breast Cancer: Cells/Tissues/Types
Speaker: Charlie Wilson, DSU
Description: The anatomy of the breast to include glandular and stromal components, lymph nodes, and anatomical relationship to other structures of the torso was presented. The basics of several types of imaging techniques were discussed and some mammograms of normal and cancerous breast were shown. The terminology for different types of breast cancer (lobular vs ductal; in situ vs invasive) was described as well as the criteria used by pathologist to assign a tumor grade.

Date: July 15, 2009
Seminar title: Breast Cancer: Epidemiology
Speaker: Charlie Wilson, DSU
Description: This lecture looked at the risk factors associated with breast cancer, its incidence and mortality, and the role of BCRA1/2 in breast cancer development.

Date: **8/12/09**
Seminar title: Inverse free boundary problem for a reaction-diffusion model of cancer growth
Speaker: Yongzhi Xu, Department of Mathematics, University of Louisville

Description: The growth of cancer cell may be modeled by a reaction-diffusion equation with free boundary. In an earlier paper, we developed a free boundary model to describe the homogeneous growth inside a cylinder, a model mimicking the growth of ductal carcinoma in situ (DCIS). Assuming that we know the coefficients of the model, we analyzed the growth tendency of DCIS. The analysis and computation of the problem show interesting results that are similar to the patterns found in DCIS. In this talk we present some inverse problems related to the free boundary model of DCIS. Assuming that we know the solution of the free boundary problem in a section of the cylinder, along with the known initial, boundary and free boundary conditions, we consider the inverse problem of finding the coefficients and the solution in the cylinder. The motivation of this problem is to develop mathematical methods to diagnose growth tendency of DCIS from biopsy data.

Date: Aug 18, 2009
Seminar title: Screening for Breast Cancer
Speaker: Shunli Zhang, MD, Kent General Hospital, Dover, DE
Description: Breast cancer is the most common and second deadliest cancer in women. The breast cancer mortality has been decreased significantly in recent years due to the screening tests. Mammography, ultrasound, and magnetic resonance imaging are the three most commonly used screening tests. Their principles, common imaging features, and guidelines are discussed.

After October 9, 2009: After running the seminars for 5 months, we named the seminar series as "DSUPENN Breast Cancer Seminar Series". A kick-off opening ceremony was held on October 9, 2009. The kick-off was well attended by 38 faculty members and students from DSU and UPENN. The DSU acting president Dr. Claibourne Smith attended the kick-off. All other seminars are well attended by an average of 22 faculty and students each seminar.

Here is the opening ceremony program:

DSUPENN Breast Cancer Seminar Series
Opening Ceremony

Applied Mathematics Research Center, DSU
Department of Radiology, UPENN

Date: October 9, 2009 (Friday)
Location: BOA 309
Time:

3.00pm Introducing the speaker, Dr. Predrag Bakic, UPENN

Seminar talk by Dr. John Lynch from UPENN Medical School.

4.00pm Introduction to the seminar series

Fengshan Liu, Applied Mathematics Research Center, DSU

Andrew Maidment, Department of Radiology, UPENN

4.15pm Welcome speech

Claibourne Smith, Acting President, DSU

4.30pm Reception

With this funded project, we will enhance DSU breast cancer research resources by: improving our expertise in translational and clinical breast cancer research; developing methods for computing image-based biomarkers for breast cancer risk, as well as methods for biomarker analysis of risk disparity; developing a database of clinical biomarkers computed from images of minority women; refining the existing and developing novel data mining techniques to determine the relationship between risk and image-based biomarkers. The improvement will support further growth of a sustained breast cancer research program at DSU and help establish us as a mid-Atlantic center for analysis of breast cancer risk and risk disparity among minority women.

The specific objectives of this training program include: (1) extending the skills of a select cadre of DSU faculty, so that they may become accomplished, influential and competitive breast cancer researchers; (2) establishing an independent breast cancer research program at DSU by performing a joint DSU–UPENN research project focused on breast cancer risk disparity in minority populations; and (3) producing a corpus of high quality published work and develop a portfolio of independently funded research grants at DSU to support a sustained breast cancer research program.

Date: October 9, 2009 (Friday)
Location: BOA 309
Seminar Title: Cell to Cell and Cell-ECM Adhesion in Cancer
Speaker: Dr. John Lynch, UPENN Medical School
Description: Cell adhesion mechanisms are especially important for the development and function of normal epithelium in many tissues including the breast. Disruption of the normal cell-cell and cell-extracellular adhesion processes contributes to carcinogenesis by promoting cell proliferation and permitting cancer cell metastasis. We will discuss several common mechanisms by which cell adhesion processes are disrupted in carcinogenesis and how the promote cancer progression.

Following this very successful opening lecture, we held 6 bi-weekly seminars to date, with another 5 scheduled by the end of Spring 2010 semesters. All the seminars were very well attended; an average attendance was 18. Following is the list of speakers and topics presented at the seminar series.

Date: October 23, 2009
Location: ETV 131
Title: Future Trends in Breast Imaging
Speaker: Andrew D. A. Maidment, Ph.D., FAAPM, Associate Professor of Radiology, Chief, Physics Section, University of Pennsylvania
Description: Medical radiography is undergoing a revolution towards quantitative tomographic imaging. As currently practiced, quantitative imaging involves the extraction of quantifiable features from images; these features add to the clinical assessment of the severity, degree of change, or relative status of a disease or injury. The field of quantitative imaging includes the development, standardization, and optimization of anatomical, functional, and molecular imaging acquisition, data analyses, display methods, and reporting. Current research is focused on the development and validation of precise image-derived metrics (image-based biomarkers) with physiologically relevant parameters, including treatment response to interventions and clinical outcomes.
As with morphologic imaging, quantitative imaging is best performed tomographically; the removal of superimposed anatomy results in more accurate and precise quantification and localization. There are two trends in tomographic x-ray imaging. The first is the increased use of computed tomography (CT); the second is the development and implementation of tomosynthesis or limited-angle computed tomography. Given that there has been significant attention paid to the high-doses associated with traditional CT, tomosynthesis is likely to become far more prevalent in the next decade. Finally, there is a revolution in radiographic contrast

agents. Taking inspiration from nuclear medicine and optical imaging, we are seeing an increase in research into radiographic contrast agents. These developments are made possible given recent advances in nanoparticles such as designer liposomes, polymersomes and nanospheres. Both blood-pool and targeted contrast agents are under investigation.

Date: November 6, 2009
Location: ETV 131
Title: Breast Cancer Epidemiology
Speaker: Shannon Lynch, University of Pennsylvania
Description: Breast cancer epidemiology is the study of the distribution and determinants of breast cancer in human populations. In the early 1990s, breast cancer incidence and mortality rates were higher in the Northeastern United States and in California, prompting breast cancer advocates to unite and push a national breast cancer agenda. Today, incidence rates and mortality rates of breast cancer are on the decline in the U.S., mostly due to advances in the prevention, screening, and treatment of breast cancer. Risk factors or determinants of breast cancer will be reviewed, as well as the role of breast cancer treatment, stage, and disease type in affecting breast cancer survival. Future directions in breast cancer research, including a focus on health disparities and environmental determinants of disease across critical periods of the human lifespan will be discussed.

Date: Dec 18, 2009
Location: ETV 131
Title: Clinical Breast Imaging
Speaker: Sara Gavenonis, University of Pennsylvania
Description: This lecture will serve as a general overview of current breast imaging technologies, including mammography, ultrasound, and breast MRI. The current discussion regarding screening mammography will also be reviewed. Future directions in breast imaging research will be covered briefly.

Date: January 21, 2010
Location: ETV 131
Title: Synopsis of Breast Pathology
Speaker: Shunli Zhang, MD, Kent General Hospital, Dover, Delaware
Description: Breast cancer is the 2nd most common malignant neoplasms in women and more than 44,000 deaths occur each year in US. Although the classification of breast neoplasms is very complicated, this lecture will cover 12 most common breast lesions, their pathological changes and the clinical significance. The pathologists' role in treating breast cancer will also be discussed.

Date: January 29, 2010
Location: ETV 131
Title: Surgical Approaches in Breast Cancer Treatment
Speaker: Julia Tchou, M.D., Ph.D., University of Pennsylvania
Description: It is an overview talking about various surgical techniques from diagnosis to treatment. I will include discussion of breast conserving surgery vs. mastectomy and sentinel node biopsy in the talk.

Date: Feb 19, 2010
Location: ETV 131
Title: The Pathology of Breast Cancer
Speaker: Carolyn Mies, MD, Associate Professor, Department of Pathology and Laboratory Medicine, University of Pennsylvania
Description: The presentation is designed to introduce the basic vocabulary & histology of human breast cancer & to describe the pathologist's role in estimating prognosis & guiding treatment.

Date: March 5, 2010
Location: ETV131
Title: Genetic Counseling and Testing for BRCA1 and BRCA2 Mutations in African American Women
Speaker: Chanita Hughes Halbert, University of Pennsylvania
Description: This presentation will describe research that is being conducted to improve decision-making about genetic testing for inherited breast cancer risk among African American women at increased risk for hereditary disease. Research on psychological and behavioral outcomes following genetic counseling will also be discussed.

Date: March 26 2010
Location: ETV131
Title: Treatment of Breast Cancer- The Medical Oncologist's Approach Multimodality
Speaker: Keerthi Gogineni, University of Pennsylvania
Description: I will be talking about the fundamentals of breast cancer, with a focus on the medical treatment of the disease.

Date: April 9, 2010
Location: ETV131
Title: Breast Cancer Radiation Therapy Treatment Techniques
Speaker: Timothy Zhu, University of Pennsylvania
Description: We will present techniques typically used for radiation therapy for breast cancer, including conventional techniques, IMRT, brachytherapy, and other special techniques.

Date: April 23, 2010
Location: ETV131
Title: Computer-aided diagnosis in mammography: from the desktop to the clinic
Speaker: Robert Nishikawa, The University of Chicago
Description: Computer-aided diagnosis (CAD) for mammography started over 25 years ago and over the years has been developed into a commercial product that is used routinely clinically. In this talk I will describe the development of CAD and the current evidence of its clinical effectiveness.

Date: May 7, 2010
Location: ETV131
Title: Multimodality Breast Imaging Biomarkers for Cancer Risk Estimation and Personalized Screening
Invited Speaker: Despina Kontos, University of Pennsylvania

Description: Growing evidence suggests that increased parenchymal pattern complexity is associated with a higher risk for developing breast cancer. Currently, the most widely used methods to quantify parenchymal complexity rely on semi-automated techniques that estimate the percent of the dense tissue in mammograms. Although useful for breast cancer risk estimation, these methods are highly subjective and difficult to standardize, potentially limiting their applicability to the general population. Emerging tomographic breast imaging modalities offer the opportunity to develop novel imaging biomarkers for quantifying parenchymal pattern complexity that may ultimately result in more accurate measures to estimate breast cancer risk. This lecture will provide an overview of emerging techniques to perform parenchymal pattern analysis using imaging modalities such as digital breast tomosynthesis (DBT), magnetic-resonance imaging (MRI) and breast ultrasound (US). Improving breast cancer risk estimation using multimodality imaging biomarkers could be of

great clinical advantage for offering customizing screening recommendations, tailoring individual treatments, and forming preventive strategies, for women at a higher risk of breast cancer.

Location: ETV131
Title: Review of Breast Cancer Basics Multimodality
Speaker: Sara C. Gavenonis, MD, University of Pennsylvania
Description: This lecture is an overview of topics covered during the 2009-2010 seminar series. The clinical "world" of breast cancer will be broadly reviewed, including epidemiology, anatomy, pathology, imaging, and current clinical oncology. Current and future research directions will also be covered, with focus on breast imaging research.

Date: January 17, 2013
Location: ETV131
Title: Task-based strategy for optimized contrast enhanced breast imaging: Analysis of six imaging techniques for mammography and tomosynthesis.
Speaker: Lynda Ikejimba, PhD student, Medical Physics Graduate Program, Duke University.
Description: Digital breast tomosynthesis (DBT) is a novel x-ray imaging technique that provides 3D structural information of the breast. Compared to 2D mammography, DBT minimizes tissue overlap potentially improving cancer detection and reducing number of unnecessary recalls. The addition of a contrast agent to DBT and mammography for lesion enhancement has the benefit of providing functional information of a lesion, as lesion contrast uptake and washout patterns may help differentiate between benign and malignant tumors. This study used a task-based method to determine the optimal imaging approach by analyzing six imaging paradigms in terms of their ability to resolve iodine at a given dose: contrast enhanced mammography and tomosynthesis, temporal subtraction mammography and tomosynthesis, and dual energy subtraction mammography and tomosynthesis.

2.1.3 Additional training activities (seminars at UPENN, mentoring sessions, and communication)

Fengshan Liu, Xiquan Shi, Charlie Wilson and Dragoljub Pokrajac attended the following UPENN Seminar.

Title: Breaking Barriers: Caring for the Underserved and Undocumented

Objectives:
· 	*Discuss the limitations to healthcare encountered by a migrant population*
· 	*Understand the value of outreach in the development of the physician*
· 	*Describe our experience with providing healthcare to a migrant population*

Jack Ludmir, MD
Professor and Chair, Department of Obstetrics and Gynecology, Pennsylvania Hospital
Vice Chairman, Department of Obstetrics and Gynecology
Director of Obstetrical Services, University of Pennsylvania School of Medicine
President of Women and Children's Health Services

Date: 	January 19, 2011 (Wednesday)
Time: 	12:00 - 1:00 PM
Location: 	Seminar Room 253, BRB II/III (Biomedical Research Building 421 Curie Blvd.) UPENN

o **Augment the faculty training by frequent communications with collaborating mentors and other renowned breast cancer researchers, by: (Y1-Y4)**

DSU faculty traveled to UPENN to attend the above-mentioned seminars, and to meet with the collaborating mentors on the same trips. Meetings are always scheduled after the seminars at UPENN. UPENN collaborating mentors, particularly Andrew Maidment and Predrag Bakic come to DSU to organize and attend the DSUPENN Breast Cancer Seminar Series. Communications are made between the seminar speakers, the mentors and DSU faculty.

- In the period January-June 2011, Among UPENN-DSU collaborative activities, Drs. Pokrajac, Maidment and Bakic supervised two Penn graduate students during their course project for the CIS537 class on Biomedical Image Analysis. The students have been working on the quantification and characterization of simulated phantom shapes using geometrical methods. The obtained results have been published in the 2012 SPIE Medical Imaging conference paper (see Chapter 4).

- Fengshan Liu attended the 2009 Annual Health Research Conference Oct. 13-14, Dover, DE on health disparity.

- On August 2-5, 2011, Fengshan Liu and Charlie Wilson attended and presented at the Department of Defense Breast Cancer Research Program "Era of Hope" Conference in Orlando, FL. (https://cdmrpcures.org/ocs/index.php/eoh/eoh2011)
  - o The poster presentation entitled "Image Based Biomarkers of Breast Cancer Risk: Analysis of Risk Disparity Among Minority Populations" (co-authored by Liu, F., Bakic, P.R., Pokrajac, D., Wilson, C., Shi, X., Kontos, D., and Maidment, A.D.A.) summarized the training and research activities performed to date on our DoD HBCU/MI PTA funded project.

- On October 6, 2011, Dr. Pokrajac presented at the Faculty of Sciences of the University of Nis, Serbia.
  - o The presentation entitled "Recursive Partitioning for Simulation of Breast Tissue," (co-authored by Pokrajac, D. and Bakic, P.R.) described the current results in simulation of breast anatomy based upon the use of recursive partitioning.

- On October 19, 2011, DSU collaborators attended the semiannual Research Retreat at UPENN
  - o The agenda included a review of the research activities within the previous 6 months, and discussion of the future activities, including the (i) Analysis of the ACRIN data; (ii) Preparation of journal and conference publications from the DoD funded project; (iii) Potential future grant proposals motivated by the results from this project; and (iv) Discussion of the simulation of partial volumes in the software breast phantoms.

- On October 21, 2011, Dr. Pokrajac presented at the Medical Image Processing Group Seminar series at the University of Pennsylvania, Philadelphia, PA.
  - o The presentation entitled "Novel Algorithm for Breast Anatomy Simulation Optimized for Generation of High Resolution Software Phantoms," (co-authored by Pokrajac, D., Maidment, A.D.A., and Bakic, P.R.) focused on the new method for generation of software breast phantoms.

- On November 4, 2011, Mr. Feiyu Chen, a Ph.D. student at DSU, presented at the Mid-Atlantic Numerical Analysis Day, organized at Temple University, Philadelphia, PA.
  - The presentation entitled "Partial Volume Simulation in Software Breast Phantoms", (co-authored by Chen, F., Pokrajac, D., Shi, X., Liu, F., Maidment, A.D.A., and Bakic, P.R.) described the initial results in developing a partial volume representation of the breast anatomy simulation.

- On December 2, 2012, Dr. Pokrajac attended a Medical Image Processing Group Seminar "Application of Multifractal Analysis in Signal and Image Processing" at the University of Pennsylvania, Philadelphia, PA, presented by Dr. Branimir Reljin, the University of Belgrade, Serbia.
  - The presentation described the results of using multifractal analysis in various signal and image processing applications, including the analysis of digitized mammogram images.

- On December 27, 2011, Dr. Pokrajac presented at the Faculty of Electrical Engineering of the University of Nis, Serbia.
  The presentation entitled "Simulation of Breast Tissue using Recursive Partitioning Algorithm," (co-authored by Pokrajac, D. and Bakic, P.R.) described the new method for simulation of breast anatomy.

- On January 27, 2012, Drs. Xiquan Shi, Fengshan Liu, and Dragoljub Pokrajac attended the Abramson Cancer Center Breast Cancer Program Scientific Retreat in Philadelphia, PA.
  The retreat agenda included the review of ongoing research within the Breast Cancer Program in the areas of basic science, breast cancer imaging, cancer risk assessment and prevention, and clinical/translational trials. Discussed were also the future collaborations and program planning related to the breast cancer survivorship, breast cancer Biobank, and other programs, meetings, and presentations.

- On February 4-9, 2012, Dr. Pokrajac and Mr. Chen attended and presented the Medical Imaging conference by the Society of Photo-Optical Instrumentation Engineers (SPIE) in San Diego, CA.
  The first presentation entitled "Partial Volume Simulation in Software Breast Phantoms" (co-authored by Chen, F., Pokrajac, D., Shi, X., Liu, F., Maidment, A.D.A., and Bakic, P.R.) described the development and initial testing of the novel method for partial volume simulation in software breast phantoms.

  The second presentation entitled "Roadmap for Efficient Parallelization of Breast Anatomy Simulation" (co-authored by Chui, J.H., Pokrajac, D.D., Maidment, A.D.A., Bakic, P.R.) described the current results in parallel implementation of the method for breast anatomy simulation.

  The third presentation entitled "Shape Analysis of Simulated Breast Anatomical Structures", (co-authored by Contijoch, F., Lynch, J., Pokrajac, D.D., Maidment, A.D.A., and Bakic, P.R.) described the development of a method for fitting ellipsoids into the simulated breast anatomical structures in the software breast phantom.

  These 3 presentations were also published in the Physics of Medical Imaging, Vol. 8313, edited by N.J. Pelc, R.M. Nishikawa, SPIE: Bellingham, WA, 2012. The published papers were included as appendices to this report.

- On March 5, 2012, Dr. Pokrajac presented at the University of Belgrade, Serbia.

The presentation entitled "Breast Tissue Simulation with Recursive Partitioning Algorithm – Latest results," (co-authored by Pokrajac, D. and Bakic, P.R.) described the new method for simulation of breast anatomy.

- On March 30, 2012, Dr. Pokrajac presented at the Faculty of Electrical Engineering of the University of Nis, Serbia – as a part of the Science Fair (sponsored by the U.S. Embassy in Serbia).
  The presentation entitled "Mathematics in the Chest" presented in a more popular manner the current results in breast tissue simulation.

- On July 8-11, 2012, Dr. Pokrajac and Mr. Chen attended and presented the IWDM 2012, the 11th International Workshop on Breast Imaging, in Philadelphia, PA.
  The first presentation entitled "Toward Breast Anatomy Simulation using GPU" (co-authored by J. Chu, D. Pokrajac, A. D. Maidment, P. Bakic) described the development of highly-parallel GPU simulation of software breast phantoms.

  The second presentation entitled "Simulation of Three Materials Partial Volume Averaging in a Software Breast Phantom" (co-authored by F. Chen, D. Pokrajac, X. Shi, F. Liu, A. D. Maidment, P. Bakic) described the development of simulation for voxels containing three different materials in a software phantom.

  These two presentations are also published in Springer Lecture Notes in Computer Science, 7361, edited by Gavenonis, Sara, Bakic, Predrag R. and Maidment, Andrew D.A., 2012.

- On February 9-14, 2013, Dr. Pokrajac attended and presented the 2013 SPIE Medical Imaging conference in Orlando, FL.
  The first presentation entitled "Breast image registration by using non-linear local affine transformation" (co-authored by F. Chen, P. Zheng, P. Xu, D. Pokrajac, P. R. Bakic, Andrew D. A. Maidment, F. Liu, X. Shi) discusses a novel method for registration of mammograms.

  The second presentation entitled "Two methods for simulation of dense tissue distribution in software breast phantoms," (co-authored by J. Chui, R. Zeng, D. Pokrajac, S. Park, K. J. Myers, A. D. A. Maidment, P. R. Bakic) described the comparison of two techniques for simulation of dense tissue distribution with the distribution from clinical images.

  These two presentations are also published in **Proceedings of SPIE Volume 8668**, *2013.*

2.1.4 Validate success of the faculty training program by semi-annual Mentorship Committee meetings for each DSU faculty, and annual teleconferences with and bi-annual visits by external Advisory Committee.

On Wednesday, Nov. 3, 2010 at 3:30-4:30pm, a teleconference meeting of the DoD award Advisory Committee was organized by Drs. Maidment, Liu, and Bakic, and attended by all the DSU faculty supported on the grant, as well as Drs. Chanita Hughes and Timothy Rebbeck from UPENN. The discussed issues include our progress on the grant, future research steps related to the genetic analysis project aims, as well as the long term aim of establishing a regional Breast Cancer Disparity Center at DSU.

DSU faculty met with UPENN mentors on January 24, 2011 and August 9, 2010 to discuss the progress and the future work of each DSU faculty.

DSU Faculty Technical Trainings at UPENN by Dr. Despina Kontos and Dr. Yuanjie Zheng
Drs. Pokrajac, Shi, and Wilson, and Ms. Fatima Boukari, a M.Sc. student at DSU, attended several training sessions on the use of the pipeline for the estimation of breast density and parenchymal texture from the ACRIN database of mammographic images. The training sessions included:

       07/20/2011: Tutorial on the pipeline
       07/29/2011: Training on the pipeline, and presentation and overview of ACRIN database
       11/16/2012: Training on pipeline code, focus on breast density
       02/24/2012: Extensive pipeline training, focus on parenchymal texture
       03/06/2012: Training on MIRC database
       03/09/2012: Additional training on MIRC database

- In August 2013, DSU graduate student, Dr. Feiyu Chen, defended his Ph.D. dissertation, "Simulation of Multimaterial Voxels in Medical Imaging Software Phantoms," based upon the research funded by this grant. Dr. Chen's committee included Drs. Shi, Pokrajac, and Bakic. The results from his dissertation were published in two conference and one journal paper (see Section 4).

- In August 2013, Dr. Bakic presented an invited talk on the Real Time Simulation of Breast Anatomy, at the 2013 AAPM Imaging Symposium on the Virtual Tools for the Validation of 3D/4D X-ray Breast Imaging, held in Indianapolis, IN. The related abstract was published in the July 2013 issue of Medical Physics. The presentation included our joint research results on improving the realism accelerating the breast anatomy simulation. The presentation also featured the activities of the AAPM Taskgroup TG234 on the Virtual Tools for Breast Imaging Validation. (Drs. Pokrajac and Bakic are members of TG234.)

- In December 2013, Dr. Maidment presented an Education Exhibit on the Role of Virtual Clinical Trials in Preclinical Testing of Breast Imaging Systems, at the 2013 RSNA Annual Meeting in Chicago, IL. At the same meeting, Dr. Bakic presented a scientific paper on the Automated and Optimized Software Platform for Virtual Clinical Trials. Both presentations were co-authored by Dr. Pokrajac, and featured results of our joint research on the breast imaging simulation.

- In February 2014, Drs. Pokrajac and Bakic attended the 2014 SPIE Medical Imaging conference in San Diego, CA to present a scientific paper on the Automated Simulation of Microcalcification Clusters in Software Breast Phantoms, as well as a Computer Demo of the Software Pipeline for Breast Imaging Simulation.

- Dr. Pokrajac presented results on partial volume simulation and mathematical issues of the simulation algorithm on XIII Serbia Mathematical Congress, 2014. Also he gave invited talk on Faculty of Electronics, Mechanical Engineering and Ship building and at Kolarac Institution in May 2014.

2.1.5 Sabbatical Leave to UPENN

    Dr. Dragoljub Pokrajac took a Sabbatical Leave to UPENN for Fall 2011 semester to complete the submission of publications related to the development of the novel method for breast anatomy simulation, and for preparation of the grant applications to be submitted to the National Institute of Health (related to the RFA on the Continued Development of Biomedical Software (**PAR-11-028):** http://grants.nih.gov/grants/guide/pa-files/PAR-11-028.html)

2.2 Objective 2

Objective 2 (from SOW): Establish an independent breast cancer research program at DSU by performing a joint DSU/Penn research project focused on breast cancer risk disparity in minority populations

This section details our activities on the proposed research project. A summary of the findings is listed in Chapter 3, and the related journal and conference publications in Chapter 4.

During the last year we have analyzed the database of anonymized clinical images previously acquired within the NIH Program Project Grant supported clinical study at the University of Pennsylvania, focused on multimodality clinical breast imaging. The study includes mammograms, digital breast tomosynthesis (DBT) images, breast MRI images, ultrasound images, and breast PET images of the same patient. Goal of the analysis was to familiarize with standard DICOM format of medical images, and to write a Matlab code for extracting and processing metadata from images.

Based on the extracted metadata, we utilized SQL to search for cases and images with particular desired properties. As an example of the image database search task, we extracted images of patients having more than 4 mammograms (more than 2 per view) per exam. A standard mammographic exam involves 4 images (2 views for each breast). A larger number of images occur due to incorrect breast positioning, or for the breasts of larger size that needed more than one mammogram per view for complete coverage. The latter is of interest for our DoD PTA project, as the incomplete breast visualization would prevent correct estimation of breast density.

Our database search task was designed in order to investigate the prevalence of cases with large breast size preventing correct density estimation.   We performed the SQL database search of 657 patients with mammography exams, and after the visual confirmation of indicated images we found 85 (i.e., 13%) cases with large breast size. Such prevalence suggests that we should develop a strategy for calculating breast density in cases with multiple mammograms per view due to the large breast size. This strategy may include merging multiple images to produce a single mammogram (which can be used for density estimation) or developing a method for combining breast density values estimated from several mammograms of the same breast.

2.2.1 Analysis of Mammographic Images and Clinical Metadata of All Minority Women and the Age-Matched Caucasian Controls from the ACRIN DMIST Database

A University of Pennsylvania IRB approval (ADD Protocol number) was obtained during the first 3 months of the project.   After having obtained the IRB approval, we transferred ACRIN-DMIST data from MIRC and resolved issues with various batch transfers. We have used all minority images from ACRIN-DMIST data and age-matched Caucasian controls [Reference]. The total number of patients was 11106 (5,553 minorities and 5,553 Caucasians). During the transfer some of the studies from ACRIN data were reported as having 'invalid' images, related to problems with encoding in the MIRC database. Some cases did not get merged in the correct MIRC folders. Those cases were reviewed and if needed pushed manually to ensure correct uploading. In addition, the DICOM import service periodically got interrupted, which required to restart the MIRC automatic importing service. VB script problem: XML files from ACRIN Data occasionally did not get parsed correctly.

We established a web-based data center for this project. The functions of the data center include (1) allowing user to design relational database on the web by providing a GUI interface to the web databases, (2) accepting data input from user manually or from remote application programs, (3) converting data from other data source format such as CSV file which is a common format for clinical data (4) allowing user to compose various

queries to select data sets from various selection caritas, (5) displaying data as web pages and export data to Excel spreadsheets for performing further data analysis.

By merging source data, converted data and computed data together to create a relational database and developing facilitating functions, we can easily implement various kinds of data selection requirements in terms of SQL statements, execute them to retrieve datasets to apply further statistics routines to discover more hidden correlations among different quantifies. For the "big data" era, such a web database is in particular helpful for clinic data analytics although it also can be used for generic purposes, for example, it can be used to teach a college-level Database Management System course.

Figure 1: The overview of all database tables used for the project done by using the Web Database Design Utility

Figure 2: Information extracted or/and computed from Dicom images is kept in this 228-field table

Figure 3: The web service utilities for managing web database and converting clinic data to relational database tables
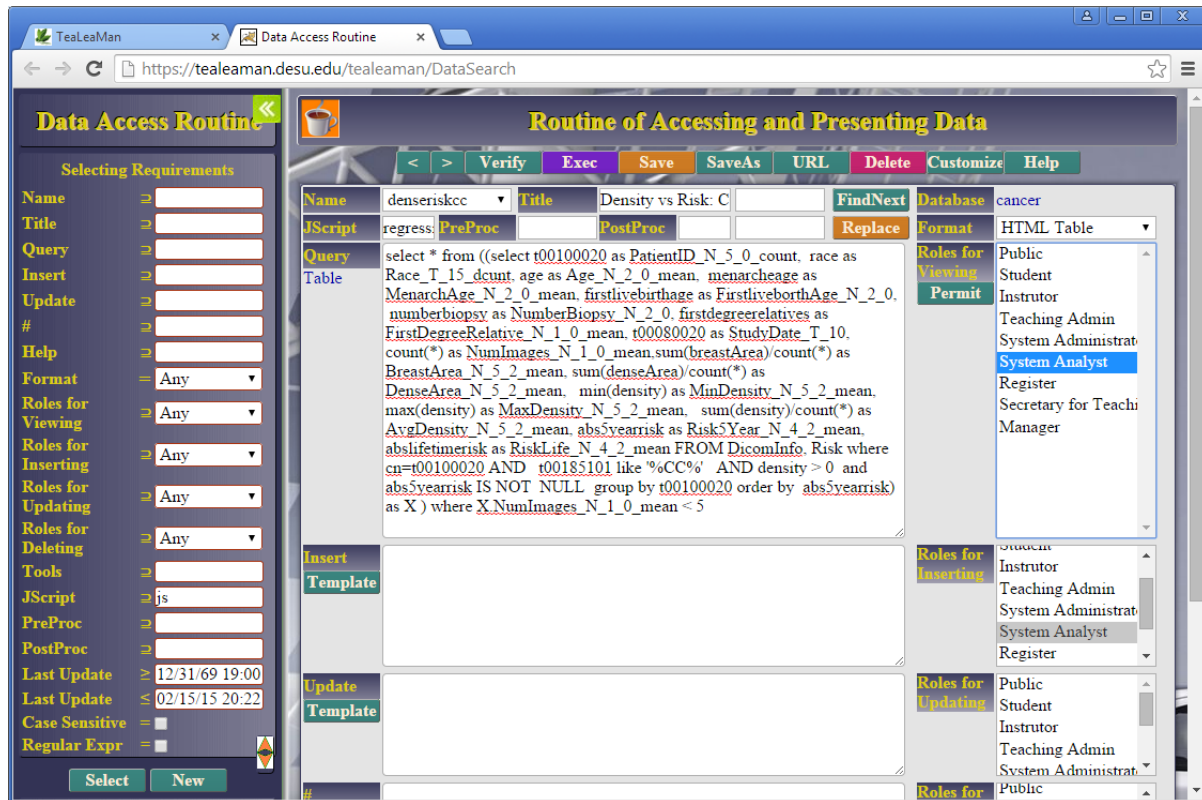
Figure 4: The Routines for Accessing and Presenting Data for composing queries to implement various data selection schemes.
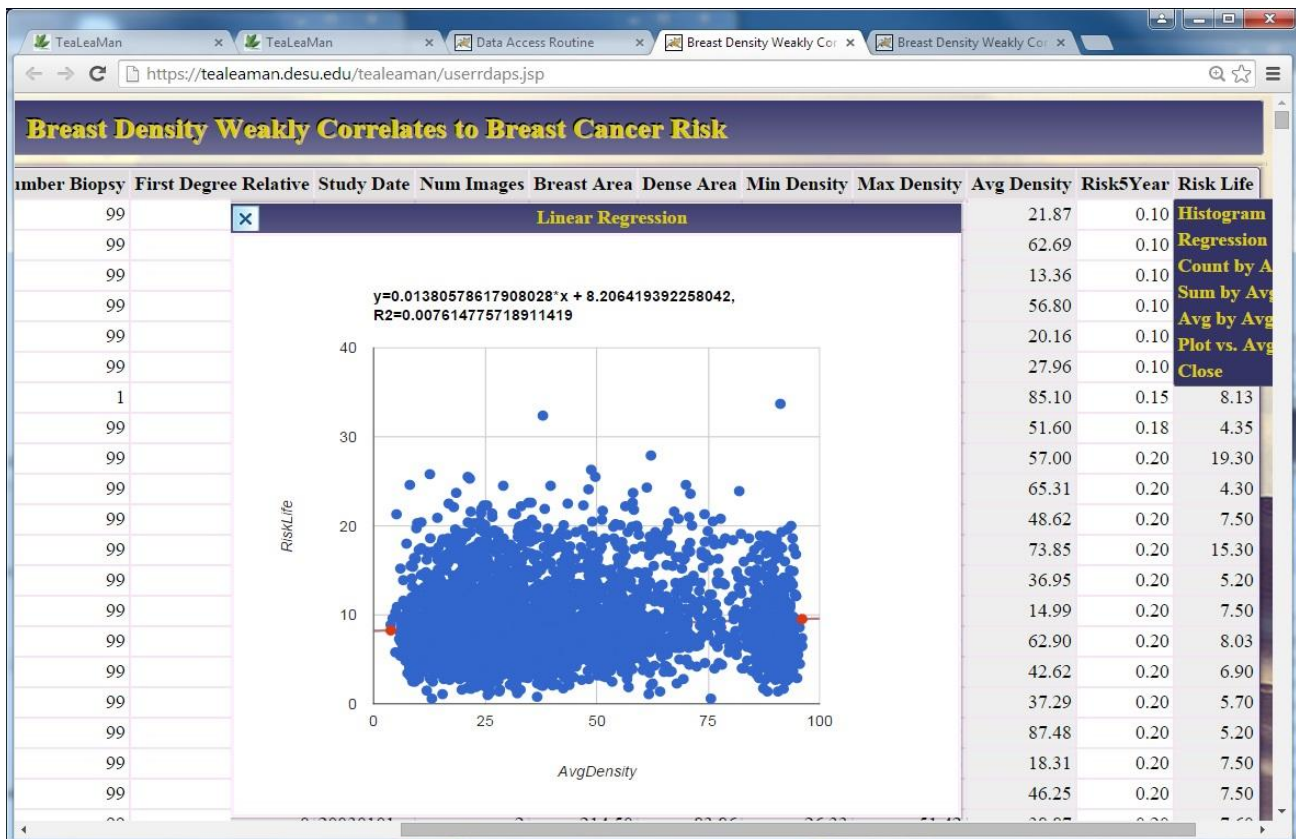
Figure 5: The Rapid Web-based Correlation Analysis for exploring various possible correlations between random variables. In this example, to explore the correlation between Average Density and Life-time Risk, click the first cell of the two columns to trig the context menu, and select Regression from the context menu.

We obtained 11106 records from the ACRIN 6652 (DMIST) study with digital images. The data include records for a sample of 5,553 minority participants and the equal number of Caucasian participants randomly drawn from the pool of age-matched controls.

We have developed a code to estimate the breast cancer risks using the demographic metadata information from the ACRIN-DMIST cases. The risk estimation is based upon the Gail risk model, currently used by the National Cancer Institute. Our code has been developed as a wrapper script around the Java software for the Gail risk estimation (downloaded from the hughesriskapps.net). The risk estimation method uses as input the patient's age, the age of the menarche, the age of the first live-birth, the number of biopsies, the number of first-degree relatives with cancer; the information about previous biopsies with hyperplasia, and the race.

This corresponding information was read from the metadata accompanying the ACRIN database of images. In cases of missing information, we followed the instructions in the Gail risk model and used the "UNKNOWN" entry.

We performed a preliminary query of the ACRIN-DMIST data aimed at identifying the prevalence of women with incomplete visualization of the breast. Here is the summary of the results from this query:
- The total number of uploaded cases with more than 4 Dicoms images is 3845; 267 of them are aggregated duplicates (files belong to the same patient but were divided in various folders);
- The number of cases with multiple images (>4) and no aggregation is 3578;
- About 30% of the cases have been checked manually to confirm partial visualization;
- If the number of DICOMs is less than 9, partial visualization is present in about 10% of the cases;
- Our estimation is that that about 500-550 out of 3845 cases have partial breast visualization which is approximately 5-8% of all cases.
- There are 4406 patients who have four or less DICOM images with MLO-position.

We developed an array of quantitative image analytics for breast density estimation and parenchymal texture analysis in digital mammography (DM) and digital breast tomosynthesis (DBT) images. These tools have been validated both in screening and diagnostic populations. Our studies have pioneered the investigation of DBT texture analysis, indicating that DBT texture features are more informative than DM texture features in characterizing parenchymal pattern. These computational are used to perform the image analysis work involved for hypothesis testing in the specific aims proposed in this project.

Breast density analysis: We developed a fully-automated breast percent density (PD%) estimation technique based on an adaptive multi-class fuzzy-c-means algorithm. Briefly, an edge-detection algorithm is applied to delineate the boundary of the breast and the Hough transform is applied to detect the boundary of the pectoral muscle. Following the segmentation of the breast, an adaptive multi-class fuzzy-c-means algorithm is applied to segment the glandular tissue, where classes are aggregated to the standard two-class dense vs. fatty paradigm using a logistic regression classification approach. Smoothing of the breast region is performed using a pixel-neighborhood medial filter of 5x5 pixel size, shown that this size can provide a good compromise between the noise reduction and preserving texture characteristics.

Parenchymal texture analysis: We developed a fully-automated software pipeline to perform quantitative analysis of breast tissue composition from multimodality digital breast images. The integrated image analytics consist of an initial image quality (IQ) test, in which a query/testing is performed in the DICOM header files of the digital images to validate acceptable dose and acquisition parameters (i.e., kVp, exposure, compression, etc.). Subsequently, the pipeline incorporates a preprocessing step with an option to create a regional tissue mask from which the imaging parenchymal pattern descriptors are extracted.
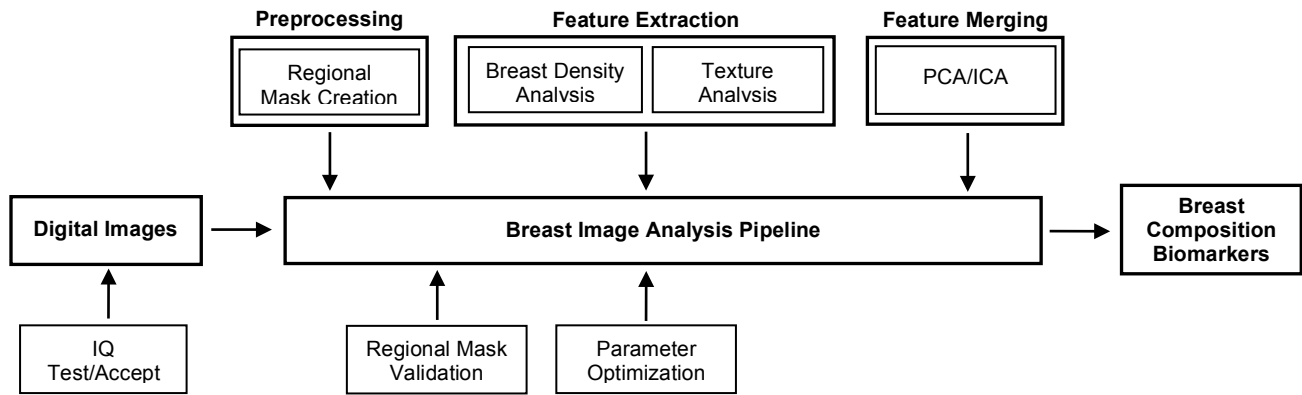
Figure 6: The flowchart of the image processing steps in our breast imaging biomarkers pipeline.

We estimated breast densities mammograms from the ACRIN database, using the software developed at the University of Pennsylvania. Note that original ACRIN-DMIST data contains images from four manufacturers. In this study, we utilized only the data from one manufacturer. The total number of processed images (both MLO and CC views) was 24945. For the same images, we calculated risk estimation using the Gail model. Racial distribution of sampled patients is shown in Table below.

Race	Number of patients
Caucasian	2496 (53.05%)
African American	1753 (37.26%)
Asian	239 (5.08%)
Hispanic/Latino	217 (4.61%)
Total minority	*2209 (46.95%)*
TOTAL	**4705 (100%)**

Table 1: Racial distribution of number of patients with estimated breast densities and breast areas

The histogram of the estimated breast density is shown in Figure below. We can see that it is most common that women's breast density is around 25%, and it is unlikely a woman's breast density is around 0%, 80% or 100%.
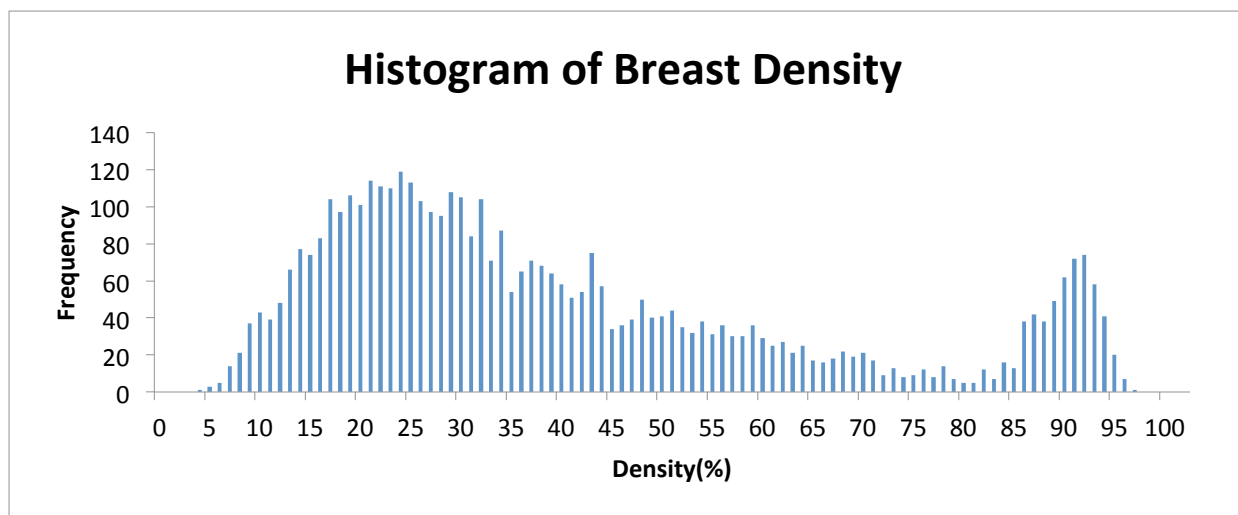
Figure 7: The histogram of breast density computed from all 4406 MLO-position images.
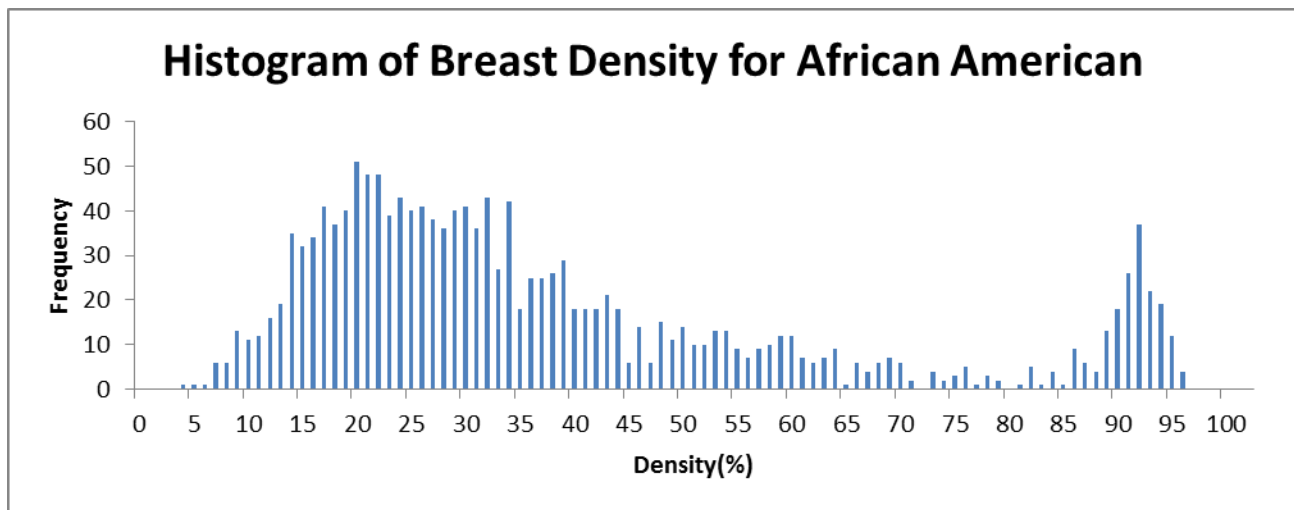
Figure 8: The histogram of breast density computed for African American.
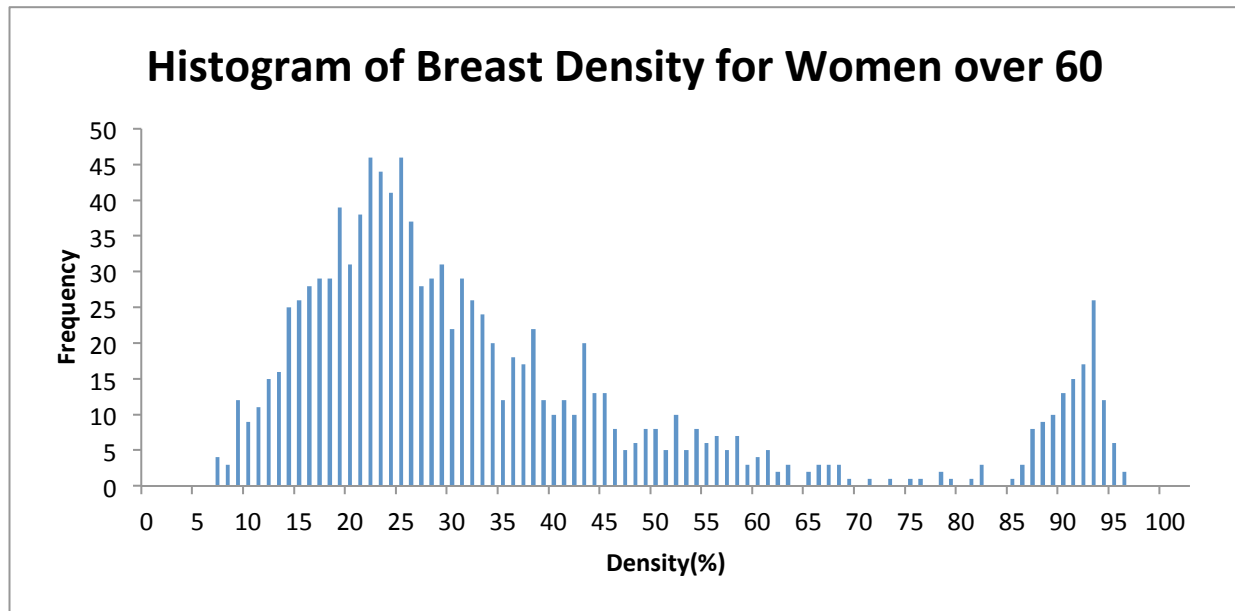
Figure 9   The histogram of breast density of women over 60 years old. Sample size=1147.

For each patient, we, for each breast (left or right) and for each view (CC or MLO) averaged the estimated breast density and breast area. We utilized the step-wise regression to determine the dependence of estimated breast density and the estimated breast area on demographics. The dependence of estimated breast density on demographic variables and breast area is significant but R^2 is very small. Race was determined as a significant variable (see figures below).
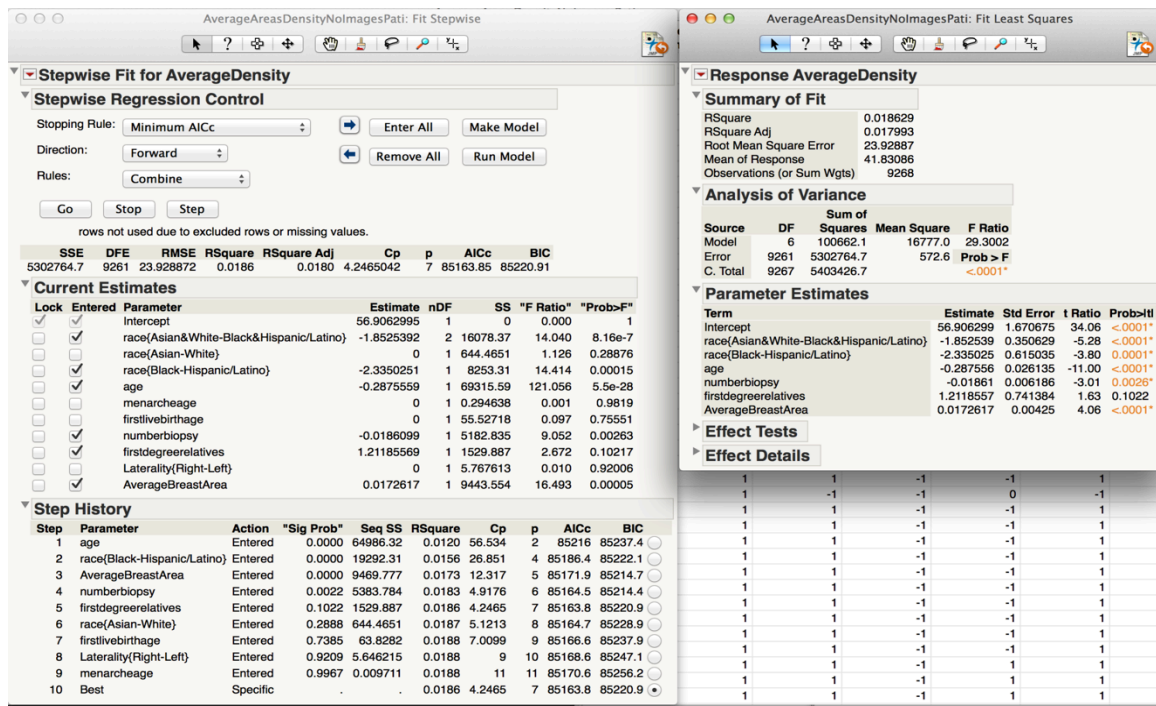
Figure 10: Step-wise regression to determine dependence of estimated breast density on race; CC view
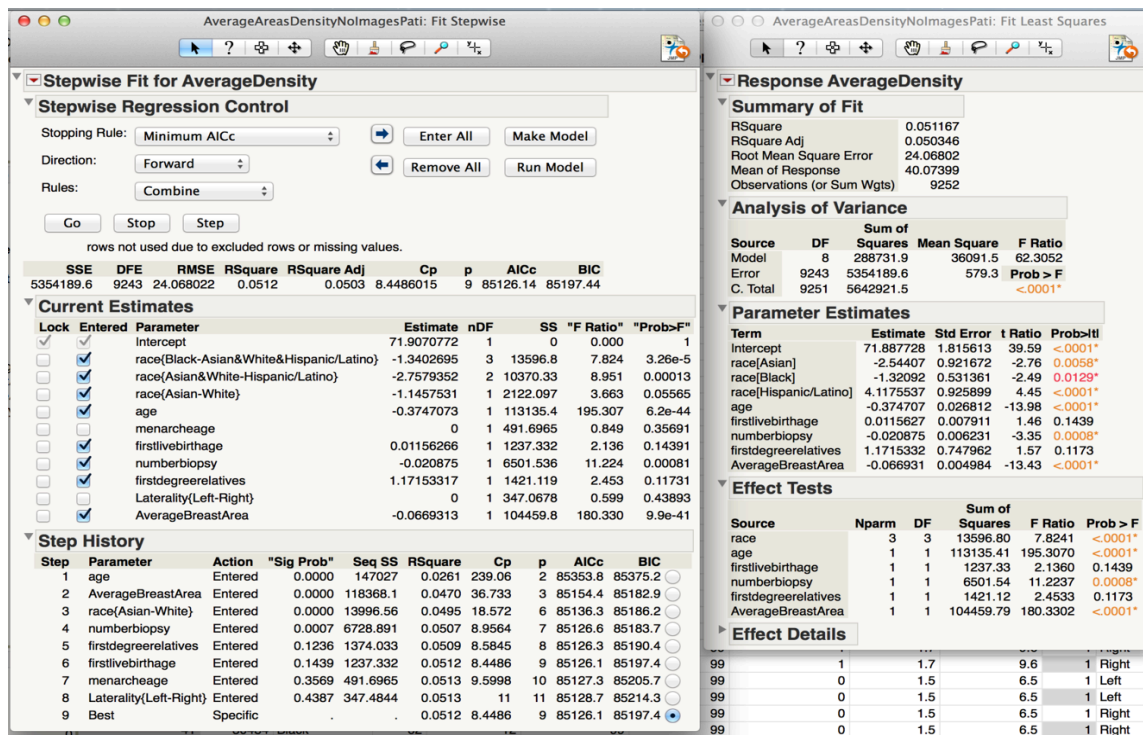
Figure 11: Step-wise regression to determine dependence of estimated breast density on race; MLO view

Step-wise linear regression also indicates that the breast area depends on race. R^2 was larger then when regressing average density. The regression model was significant (see Figures below).
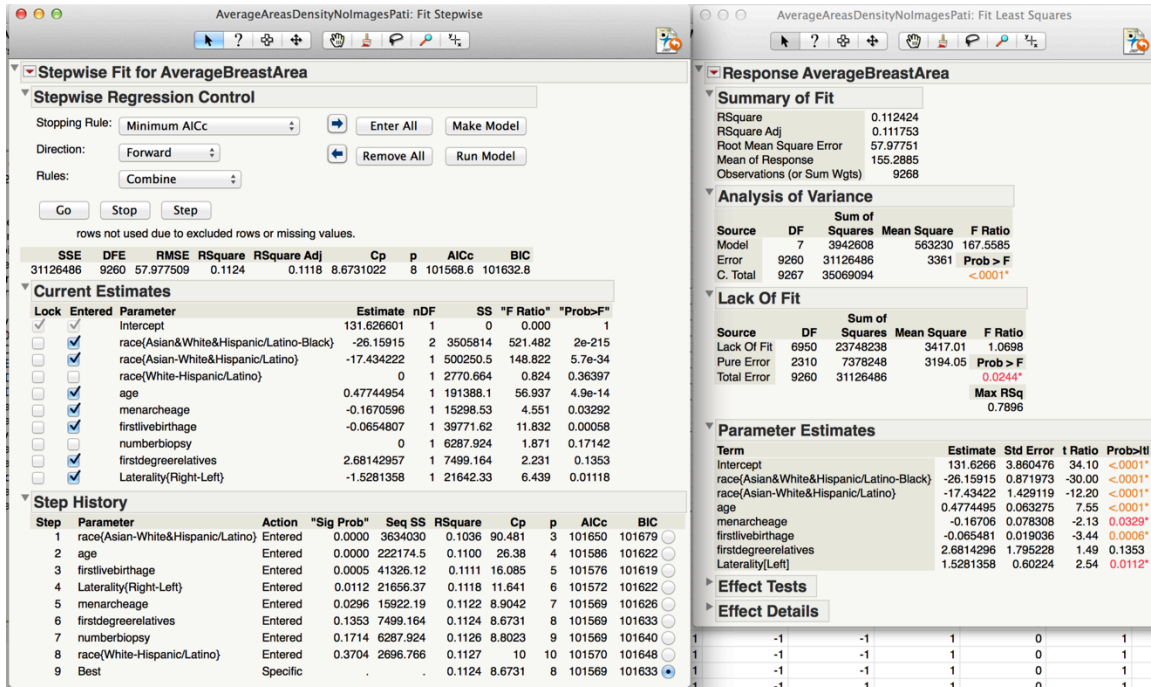
Figure 12: Step-wise regression to determine dependence of estimated breast area on race; CC view
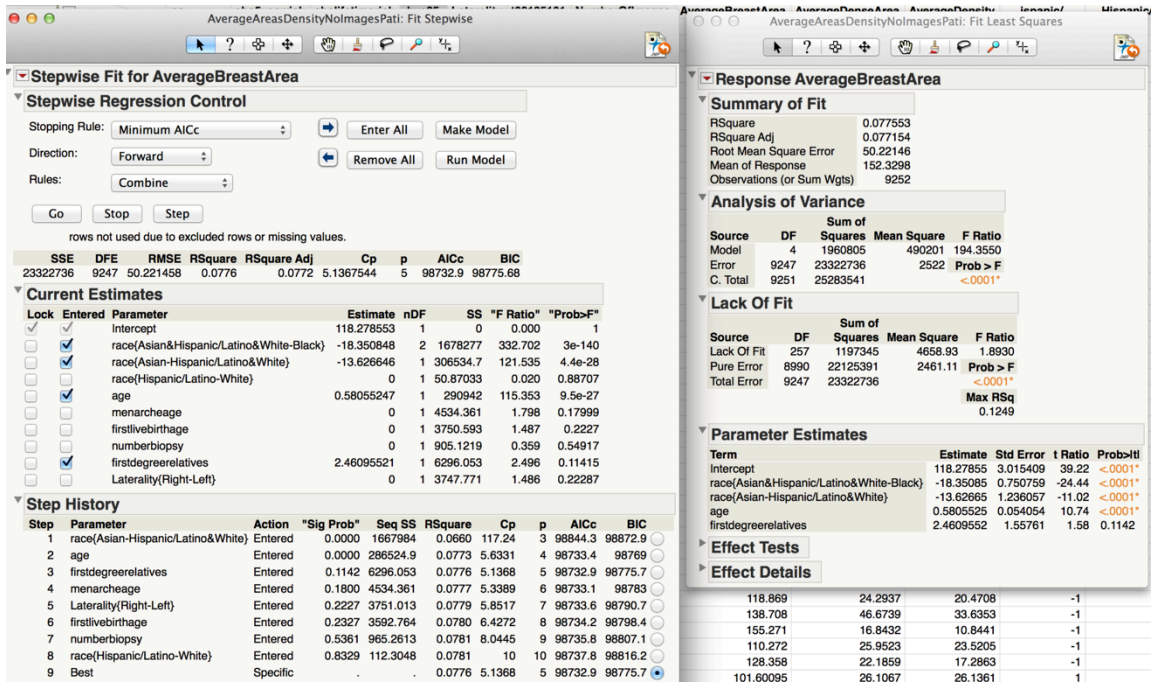
Figure 13: Step-wise regression to determine dependence of estimated breast area on race; MLO view

2.2.2 Analysis of risk prediction models

For each patient from the dataset of 4705 patients (see Subsection 2.2.1), we, for each breast (left or right) and for each view (CC or MLO) averaged the estimated breast density and breast area. Based on this data, we wanted to test the hypothesis whether the estimated GAIL risk and estimated breast density are correlated. The results, shown below for CC view, indicate the correlation to be significant but small. Our conclusion is that the breast density can be added to a risk model (as a variable orthogonal to the existing demographic variables utilized).
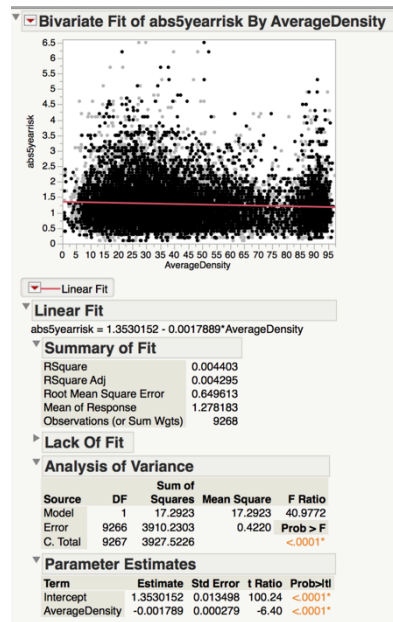
Figure 14: Correlation between average estimated breast density and 5years risk estimated by GAIL model; CC view
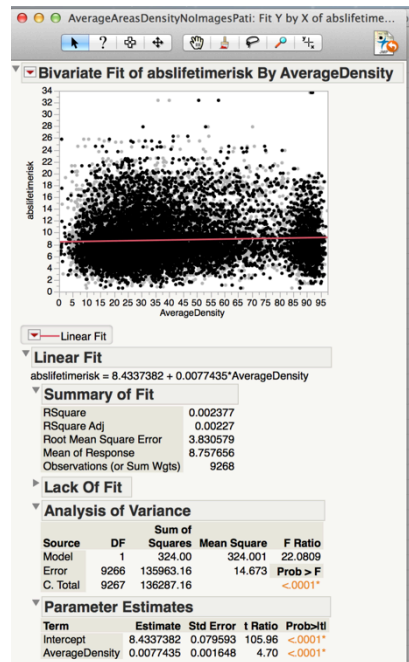
Figure 15: Correlation between average estimated breast density and lifetime estimated by GAIL model; CC view

We performed a preliminary verification of this hypothesis. From the considered dataset we extracted patients with performed biopsies (123 patients with negative biopsy results, the value of bxe25 attribute=1, and 32 patients with positive biopsies results, bxe25=2). We performed logistic regression of biopsy result on estimated GAIL risks (lifetime risk, 5 years risk), and compared the classification results when the estimated breast density and the estimated breast areas are added as independent regressors. The results shown on Figure 16 below indicate that the accuracy of classification cannot be improved by adding the estimated image parameters on CC view images. The results on MLO view images, Figure 17 indicate that the prediction model could be improved by adding the estimated *breast area* as an independent variable.

Figure 16: Logistic regression of biopsy results based on (a) GAIL lifetime and 5 years risk; (b) GAIL lifetime and 5 year risk and estimated breast density and area; 155 patients (total of 290 images with left and right orientation), CC view.
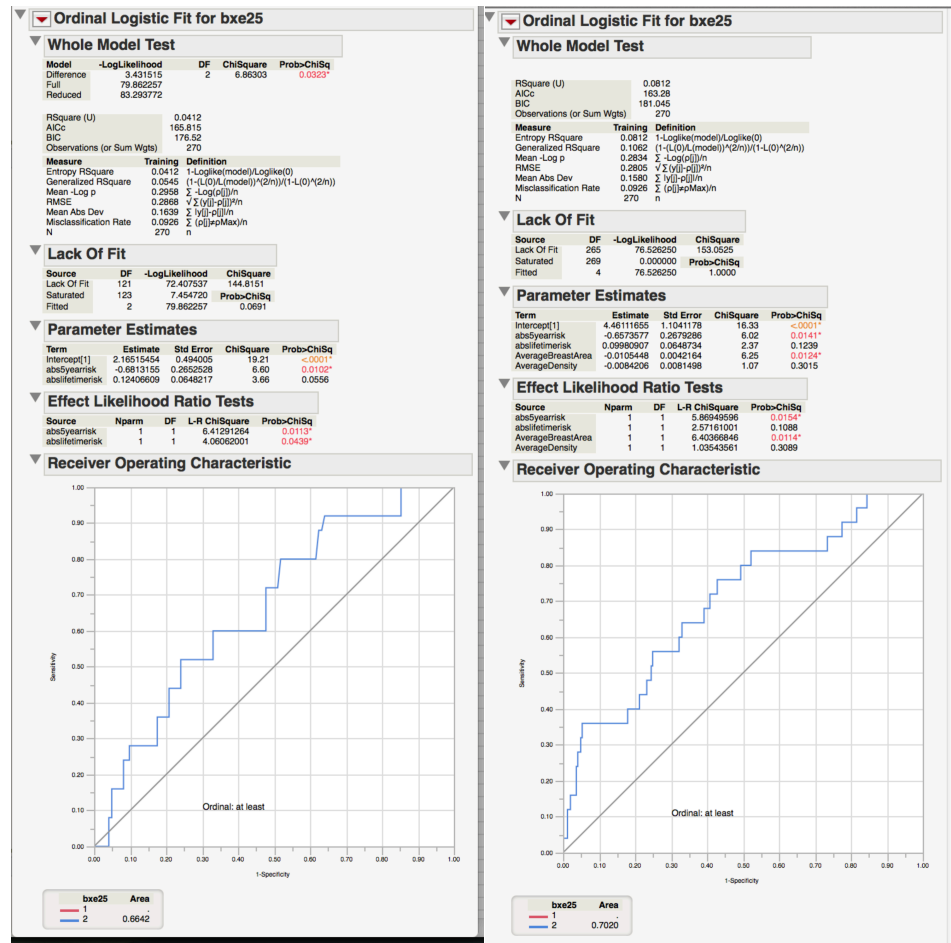
Figure 17: Logistic regression of biopsy results based on (a) GAIL lifetime and 5 years risk; (b) GAIL lifetime and 5 year risk and estimated breast density and area; 155 patients (total of 270 images with left and right orientation), MLO view.

We have explored how the validity of hypothesis depends on age. We find that in most age intervals, the breast density is positively but weakly correlated to the breast cancer risk, but in other intervals, the correlation is weakly negative, as shown in the Table 2. The linear regression coefficients indeed depend on the age heavily. In particular, as shown in Figure 18, in the age interval 40-45 and 62-72, the positive correlation between breast density and breast cancer risk is much stronger than other age intervals.

Center of Age Interval	Number of Patients	Slope of Linear Regression	Coefficients Of Determination
36	113	0.032843	0.023696
40	135	0.004559	7.67E-04
41	139	0.001818	1.41E-04
42	161	0.021103	0.014249
43	146	0.010153	0.003441
44	145	0.025072	0.023202
45	192	0.005164	0.001121
46	198	-0.0091	0.005601
47	182	-0.00683	0.001798
48	173	-0.01293	0.010095
49	176	5.00E-04	1.17E-05
50	175	0.00754	0.003063
51	210	-0.01771	0.012417
52	174	0.002875	3.99E-04
53	219	-0.00871	0.003447
54	176	-0.01311	0.007823
55	171	-5.32E-04	1.64E-05
56	158	0.016326	0.012772
57	158	-0.01571	0.01192
58	140	0.011876	0.006563
59	131	0.014717	0.011779
60	116	-0.0165	0.013812
61	118	-0.00179	2.53E-04
62	103	0.008103	0.003058
64	169	0.009127	0.006134
66	180	0.010356	0.011963
68	135	0.014086	0.018304
70	104	0.024838	0.097342
72	121	-0.00443	0.002877
76	109	-0.00644	0.010156
82	53	-0.00736	0.017216

Table 2. This table shows how much the breast density is correlated to the breast cancer risk in different age intervals
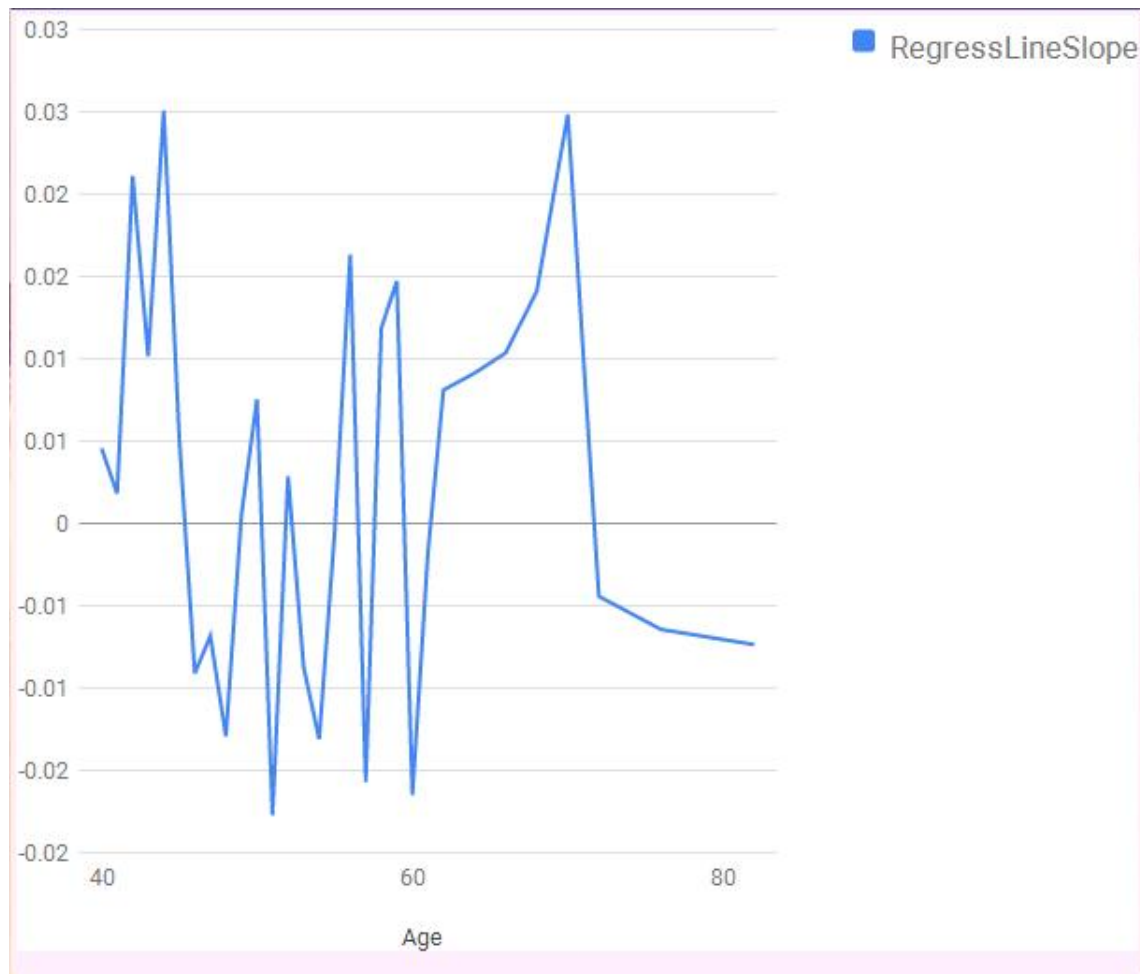
Figure 18 The slope of the linear regression varies with age. In the age interval 40-45 and 62-72, the correlation is positive.

2.2.3 Mammogram Image Registration/Fusion

During the previous review of mammographic images from the UPENN clinical ("PPG") database, we noticed that a number of women with large breasts had multiple mammographic views, each showing only a partial visualization of the breast. The number of cases in the "PPG" database with such incomplete visualization was estimated to about 13% of all images. A similar observation was made for the mammographic images from the ACRIN database used in this DoD funded research project, although with lower prevalence (5-8%). Accurate estimation of breast percent density from mammographic images requires a complete visualization of the breast tissue. For that purpose, as one of our research sub-aims, we have been developing a method for registration/fusion of multiple mammographic images of the large breasts.

Individual images of a large breast have been acquired with different x-ray tube angles and using different mammographic compression force. These variations may cause various deformations between individual images, including affine (due to the change in tube position) and non-linear (due to the elastic properties of the breast). Since the individual images capture different portions of the breast, it is needed to identify a common region of the breast tissue, visualized in both images. The next step towards the fusion of common regions in two mammograms includes detection of correspondence (or fiducial) features in the two images. These fiducial features include edges, as detected at the breast skin outline and in the breast interior in mammograms, as well as the texture features within the breast interior. We are currently in the process of selecting the appropriate features to be used to drive the registration method.

A novel breast image registration method is proposed to obtain a composite mammogram from several images with partial breast coverage, for the purpose of accurate breast density estimation. The breast percent density estimated as a fractional area occupied by fibroglandular tissue has been shown to be correlated with breast cancer risk. Some mammograms, however, do not cover the whole breast area, which makes the interpretation of breast density estimates ambiguous. One solution is to register and merge mammograms, yielding complete breast coverage. Due to elastic properties of breast tissue and differences in breast positioning and deformation during the acquisition of individual mammograms, the use of linear transformations does not seem appropriate for mammogram registration. Non-linear transformations are limited by the changes in the mammographic projections pixel intensity with different positions of the focal spot. We propose a novel method based upon non-linear local affine transformations. Our algorithm requires that feature points be extracted prior to registration, and the result of registration will depend on the reliability and accuracy of the extracted features. Automatic identification and extraction of feature points is difficult due to the non-linear compression deformation and the lack of significant landmarks in mammograms. We observe the prominent features (such as ducts and blood vessels) from both images. The crossing points are determined upon visual similarity in both mammograms. Due to compression and different positions of the breast, the coordinates of those crossing points may be different in the two mammograms, but the orientation of feature and local curvature of crossing points are more likely to be preserved. We also select other features (end points and middle points) in a small neighborhood around the selected crossing points. Subsequently, the deformation between two sets of feature points can be estimated. Given two sets of feature points in two images that need to be registered, we assume the deformation between them can be approximated by affine transformation, which can be considered as a first-order approximation of the true transformation resulting from breast projection. Finally, Shepherd interpolation is employed to compute affine transformations for the rest of the image area. The pixel values in the composite image are assigned using bilinear interpolation. We present preliminary results using the proposed approach applied to clinic mammograms taken from the ACRIN DMIST database of mammograms. This work is a part of a larger study of racial disparity in breast cancer risk. For that project, breast percent density and parenchymal texture of minority women and age-matched Caucasian controls from the ACRIN DMIST database are being compared. To date, we have been able to achieve anecdotal results that support continued development and testing of this new method. The proposed method is robust, since the results of registration are similar regardless of the choice of the reference image. The observable features, especially the nipple and the boundary of skin, have good agreement. The results of the proposed method are comparable to the results of the diffeomorphic transform implemented using ANTs, an open source software package. Particularly, the textures of warped image are preserved in registered images, and the shape of registered image is similar as reference image. The registration error is smaller in the region of overlap (the upper part of the registered image), since we can extract the corresponding feature points only from this region. . The proposed transformation can be controlled locally. Moreover, the method is converging to the ground truth deformation if the paired feature points are evenly distributed and its number is large enough .In our future work, we plan to perform more extensive quantitative validation of the proposed algorithm on a series reference and warped images extracted from all the applicable images in the ACRIN DMIST database. Also, we will apply the technique to more images in the ACRIN DMIST database and develop statistical measures of the registration accuracy.

Most recently, we have also introduced another new image registration method by using multivariate spline functions. Different from the product form splines used previously, the multivariate spline we used is non-product form, which is much more flexible in application. That is, non-product form multivariate spline functions can be conveniently used to approximate real data locally. Actually, for the corresponding portions of an overlap part of two images, one can be locally treated as the image obtained by an affine transformation of the other. This indicates that we can partition two images into small parts and then register each small part onto its corresponding portion of the reference image. This is the basic idea of our spline image registration method. To facilitate the application with previous results, the range of multivariate spline function is taken as matrix variable. The basic idea of our method is as follow: we first translate one image, called reference image, into the output image, and then register others, called source images, to the output image by spline functions. Since quadratic is the lowest possible degree to construct smooth multivariate splines, we employ quadratic multivariate spline defined over Powell-Sabin type triangulations to image registration. The experiment results

show that, comparing previous methods, the registration outputs have better matching and keep more details with better fair looking.

During this granting period Drs. Pokrajac and Bakic arranged several Skype teleconferences with the research lab of Dr. Aleksandar Peulic from the University of Kragujevac, Serbia, who continued the project previously performed by Dr. Feiyu Chen and Mr. Penglong Xu. Also, Dr. Pokrajac met Dr. Peulic and his group at his research trip to Serbia in October 2013. Currently, Dr. Peulic is working on registration of mammographic images belonging to large breasts using ANTs software. The initial results indicate the possibility of good stitching of images belonging to the same breast.

2.2.4   Breast Phantom Simulation and Analysis

We have originally proposed Exploratory study 2 on potential racial differences in genetic determinants of breast density. That exploratory study has not been performed, partly due to Charlie Wilson's health condition.

Instead, we have performed an exploratory research on refining the computer simulation of breast anatomy and imaging, previously developed at UPENN. This study has been instrumental in the development of the registration method needed in to perform fusion of mammograms with partial coverage in large breast size (Section 2.2.3). This exploratory study has received significant attention in the breast cancer research community, as evident by related publications and the grant funding proposals.

During the development of registration/fusion method (section 2.2.3), we have used simulated mammograms of the breast software phantoms developed at the University of Pennsylvania. In preparation of this task, we have worked on the breast tissue modeling, to allow for faster generation of mammograms with the resolution comparable to the clinical images. Images of the breast software phantom previously developed at the University of Pennsylvania sometimes include quantization artifacts due to the large voxel size. Increasing the phantom voxel size (above 100-200 microns) requires a prohibitively long simulation.

We have developed an efficient method for generating anthropomorphic software breast phantoms with high resolution. The present method has been optimized for computational complexity to allow for fast generation of the large number of phantoms required in virtual clinical trials of breast imaging. The new breast anatomy simulation method performs a direct calculation of the Cooper's ligaments (i.e., the borders between simulated adipose compartments). The calculation corresponds to quadratic decision boundaries of a maximum *a posteriori* classifier. The method is multiscale due to the use of octree-based recursive partitioning of the phantom olume. The method also provides user-control of the thickness of the simulated Cooper's ligaments and skin. Using the proposed method we have generated phantoms with voxel size in the range of (25-1000 μm)^3/voxel. Experimental and theoretical considerations show that the computational time increases with the square of the inverse voxel size (compared to at least cubic complexity of our previous region growing algorithm). This way we could achieve the voxel size in the order of the detector pixel size (~70μm). Note that the simulation of phantoms with voxel size of 50 μm or less was not feasible using the previous techniques.

We designed methods to evaluate the achieved thickness control of the Cooper's ligaments in the simulation algorithm. Also, we have proposed improved thickness control algorithm which is currently under development.

Further modification to our simulation algorithm is proposed, in order to improve the quality of simulated projections generated using software breast phantoms. Anthropomorphic software breast phantoms have been used for quantitative validation of breast imaging systems. Previously, we developed a novel algorithm for breast anatomy simulation, which did not account for the partial volume (PV) of various tissues in a voxel; instead, each phantom voxel was assumed to contain single tissue type. As a result, phantom projection images displayed notable artifacts near the borders between regions of different materials, particularly at the skin-air

boundary. These artifacts diminished the realism of phantom images. One solution is to simulate smaller voxels. Reducing voxel size, however, extends the phantom generation time and increases memory requirements. We achieved an improvement in image quality without reducing voxel size by the simulation of PV in voxels containing more than one simulated tissue type. The linear x-ray attenuation coefficient of each voxel is calculated by combining attenuation coefficients proportional to the voxel subvolumes occupied by the various tissues. A local planar approximation of the boundary surface is employed, and the skin volume in each voxel is computed by decomposition into simple geometric shapes. An efficient encoding scheme is proposed for the type and proportion of simulated tissues in each voxel. We illustrated the proposed methodology on phantom slices and simulated mammographic projections. Our results show that the PV simulation has improved image quality by reducing quantization artifacts. A general case for simulation of the partial volume (PV) averaging in software breast phantoms is considered. We studied results of simulated PV in a general case of voxels containing up to three materials. Local planar approximations of boundary surfaces are employed. The material proportions in each voxel are computed by decomposition into geometric shapes.

A roadmap has been proposed to optimize the simulation of breast anatomy by parallel implementation, in order to reduce the time needed to generate software breast phantoms. The rapid generation of high resolution phantoms is needed to support virtual clinical trials of breast imaging systems. The proposed roadmap for efficient parallelization includes the following steps: (i) migrate the current code to a C/C++ platform and optimize it for single-threaded implementation; (ii) modify the code to allow for multi-threaded CPU implementation; (iii) identify and migrate the code to a platform designed for multithreaded GPU implementation. As the first step of the proposed roadmap we have identified a bottleneck component in the MATLAB implementation using MATLAB's profiling tool, and created a single threaded CPU implementation of the algorithm using C/C++'s overloaded operators and standard template library. The C/C++ implementation has been compared to the MATLAB version in terms of accuracy and simulation time. A 520-fold reduction of the execution time was observed in a test of phantoms with 50-400 μm voxels. In addition, we have identified several places in the code which are to be modified to allow for the next roadmap milestone of the ultithreaded CPU implementation. Further, the development of a multithreaded GPU implementation is under way.

Recent advances in high-resolution 3D breast imaging, namely, digital breast tomosynthesis and dedicated breast CT, have enabled detailed analysis of the shape and distribution of anatomical structures in the breast. Such analysis is critically important, since the projections of breast anatomical structures make up the parenchymal pattern in clinical images which can mask the existing abnormalities or introduce false alarms; the parenchymal pattern is also correlated with the risk of cancer. As a first step towards the shape analysis of anatomical structures in the breast, we have analyzed an anthropomorphic software breast phantom. The shape analysis was performed by fitting ellipsoids to the simulated tissue compartments. The ellipsoidal semi-axes were calculated by matching the moments of inertia of each individual compartment and of an ellipsoid. The distribution of Dice coefficients, measuring volumetric overlap between the compartment and the corresponding ellipsoid, as well as the distribution of aspect ratios, measuring relative orientations of the ellipsoids, were used to characterize various classes of phantoms with qualitatively distinctive appearance. A comparison between input parameters for phantom generation and the properties of fitted ellipsoids indicated the high level of user control in the design of software breast phantoms. The proposed shape analysis could be extended to clinical breast images, and used to inform the selection of simulation parameters for improved realism. The following subsections discuss in more detail some of the achievements in breast simulation.

Development of novel phantom

We designed and implemented the new software breast phantom in Matlab and analyzed its asymptotic spatial and temporal complexity. We experimentally evaluated temporal complexity by generating more than 400 phantoms of different sizes, resolutions, thickness and shapes of the compartments. The phantoms were generated with various voxel sizes in the range of 25-10000 micrometers, see Figure below.
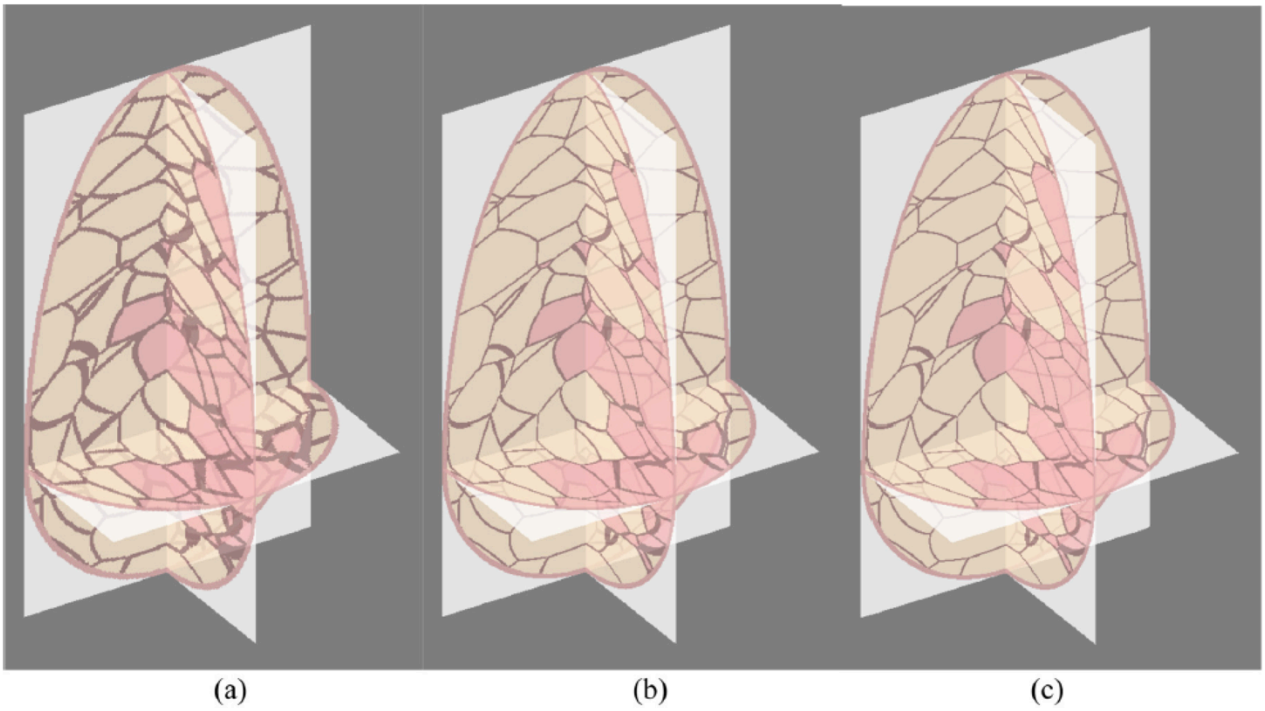
Figure 19: The same geometry of phantom simulated at (a) 400um; (b) 100um and (c) 25um resolutions (Medical Physics, 2012)

The simulation time as function of the voxel size, number of compartments and thickness was approximated using the power regression. We also performed a visual comparison of the phantoms generated at different voxel size. We experimentally demonstrated the power exponent (equal to the slope in the log-log graph) w.r.t. the voxel size of 2 in contrast to the slope larger than 3 for the old method, see Figure below.
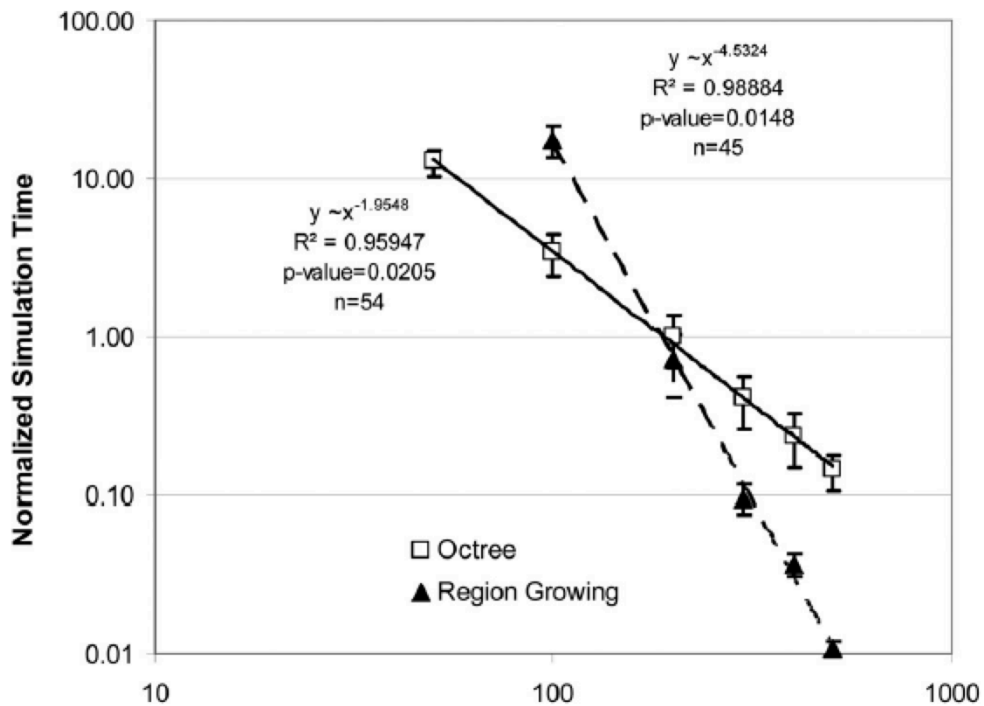
Figure 20: Experimental comparison of computational complexity of recursive partitioning and region growing algorithms (Medical Physics, 2012).

This is reflected in a progressively faster simulation for phantoms with voxel size smaller that 200 micrometers. Specifically, we were able to generate phantoms at 50um and 25um resolutions which are not feasible by region growing. Generating phantoms with voxel size below 200 micrometers is of importance for reducing quantization artifacts in simulated phantom images, since clinical x-ray detectors currently produced with the pixel pitch down to 50 micrometers. The visual comparison of the phantoms generated at different voxel size confirmed an improved quality of simulated anatomical structures, as reflected in reduced quantization artifacts, see Figure below.

(a)   (b)
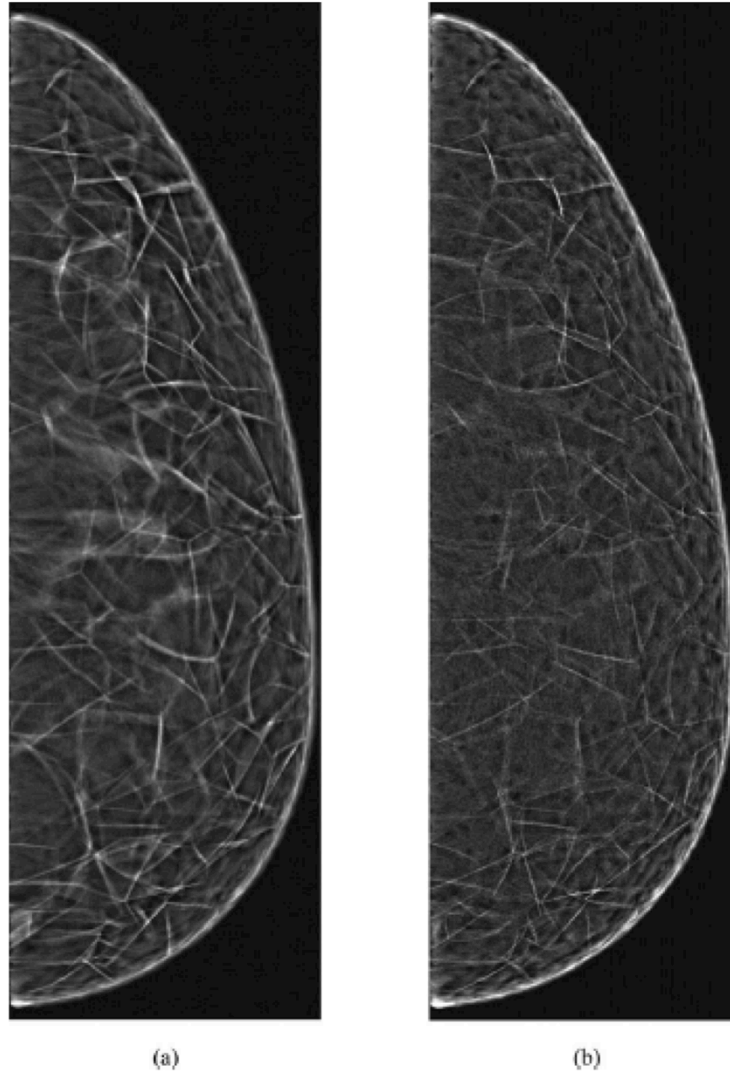
Figure 21: Synthetic mammographic projections through the compressed phantoms with voxel size of 200um and 333 compartments simulated (a) using the recursive partitioning algorithm; (b) using the region growing algorithm (Medical Physics, 2012).

We provided a formal proof that the recursive partitioning algorithm has quadratic complexity in terms of the inverse linear voxel size, and that the algorithm is computationally optimal (hence, there cannot exist an algorithm with lower asymptotic computational complexity).

Simulation of compartment microstructure

We designed an implemented a further improvement of the simulation algorithm that can also simulate compartment microstructure, see Figure below.
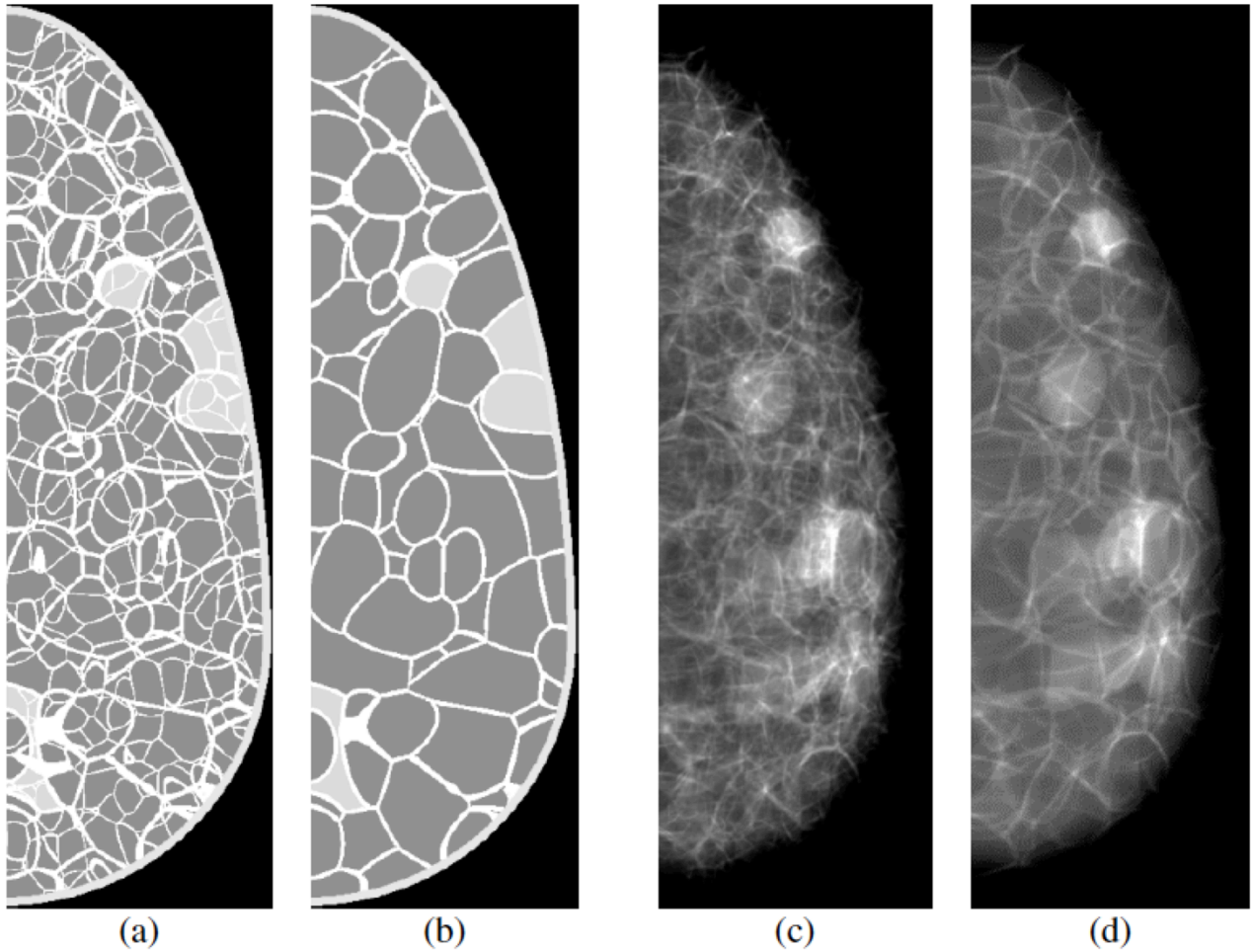
Figure 22: Simulation of tissue microstructure. Shown are sections of a software phantom (a) with and (b) without subcompartments with corresponding synthetic mammographic projections (c) and (d). (IWDM, 2014).

The initial results, see Figure below, indicate that such a novel method can improve phantom realism.
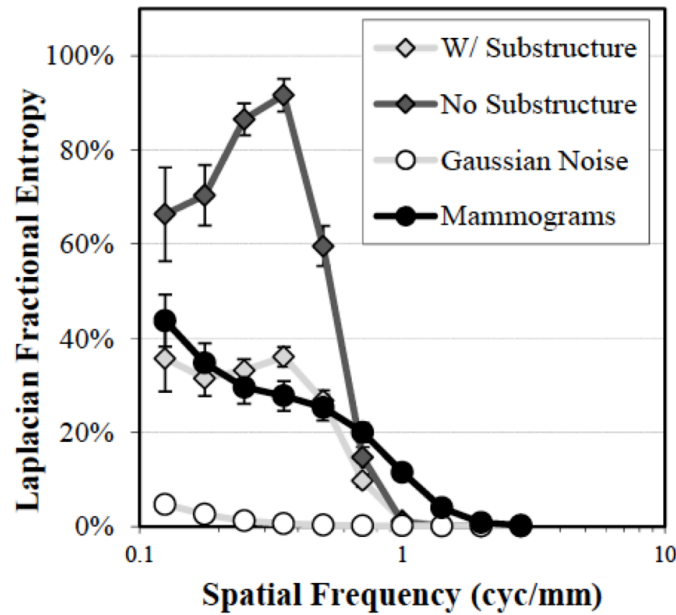
Figure 23: Laplacian fractional entropy for different versions of phantoms (SPIE, 2014).

Automated Insertion of Simulated Microcalcification Clusters into Software Breast Phantoms

We designed and implemented a method for automatic insertion of calcification clusters in the phantom and evaluate different strategies for cluster positioning, 3D clusters of microcalcifications, extracted from reconstructed clinical images, are inserted at randomly selected positions out of a set of the candidate positions. The candidate positions are identified based upon the assumptions about the origin of microcalcifications. Directed placement is based upon the assumption that clusters may occur only in non-adipose tissue regions; undirected placement presumes that clusters may be found anywhere inside the breast. In both cases, the candidate positions are identified by convolving a 3D rectangular hull (around the cluster) with phantom regions of non-adipose tissue (in case of directed placement) or with the breast interior (for undirected placement). These two placement strategies were validated in a 2-alternative forced choices observer study with 3 clinical radiologists as observers. Each observer reviewed 450 image pairs and indicated their preference between the directed and undirected placement, see Figure 20 below. The study suggested observer's preference for undirected placement. More details are found in our 2014 SPIE talk.
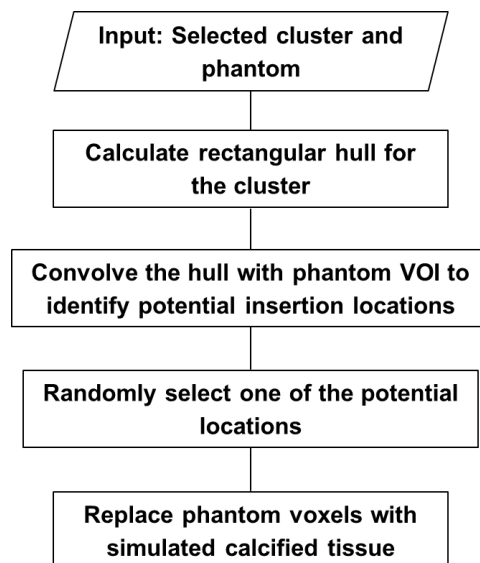
Figure 24: Flowchart of the method for automated insertion of simulated clusters into a software phantom
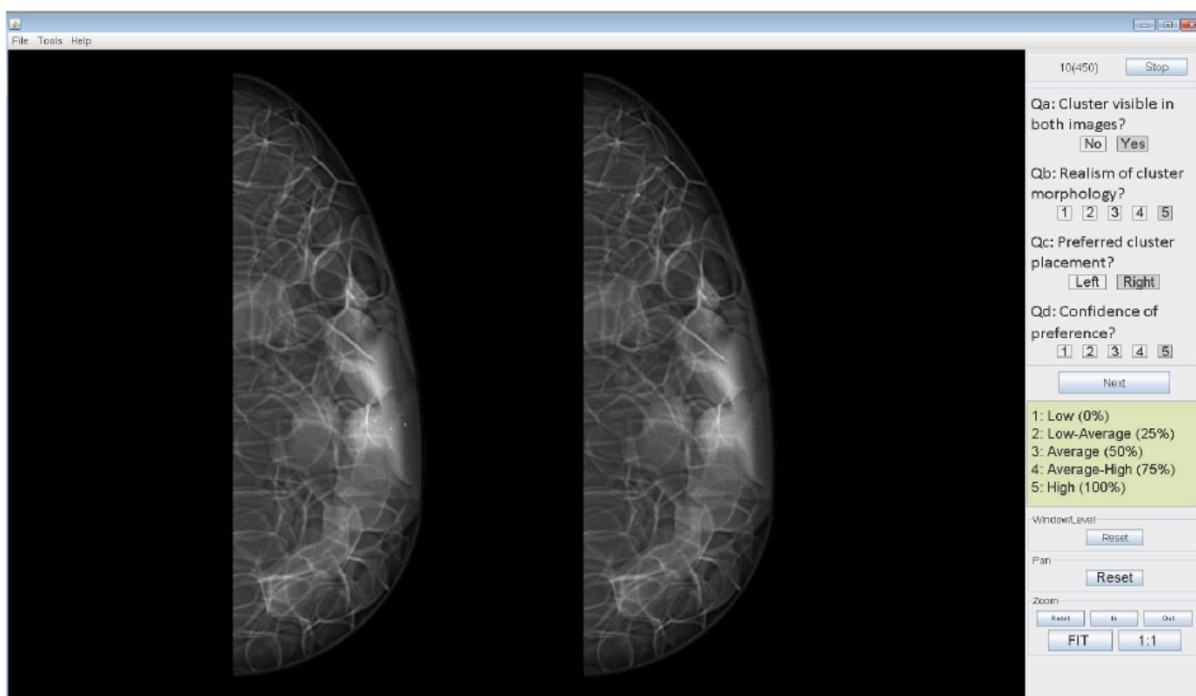
Figure 25: 2-AFC observer study user interface for evaluation of calcification cluster insertion strategies (SPIE, 2014).

Partial volume simulation

We developed a novel method to simulate voxels that contain multiple materials (e.g., skin and air. Skin and adipose tissue). We designed an efficient encoding scheme to treat different possible cases of such voxels, see Table below.

Case	p_1 (6 bits)	p_2 (6 bits)	Label (4 bits)
1. Skin	0	0	0
2. Air	0	p_{Air} (=100)	0
3. Cooper's ligament	0	0	1
4. Fat	0	0	2
5. Dense	0	0	3
6. Skin; air	0	p_{Air}	0
7. Skin; fat tissue	0	p_{Skin}	2
8. Skin; dense tissue	0	p_{Skin}	3
9. Skin; Cooper's ligament	p_{Cooper}	0	0
10. Cooper's ligament; fat	p_{Fat}	0	1
11. Cooper's ligament; dense	0	p_{Dense}	1
12. Skin, Cooper's ligament and fat tissue	p_{Cooper}	p_{Skin}	2
13. Skin, Cooper's ligament and dense tissue	p_{Cooper}	p_{Skin}	3

Table 3: Encoding partial volume voxels

For voxels containing skin and air, we used the linear approximation of skin/air boundary, as illustrate in Figure below.
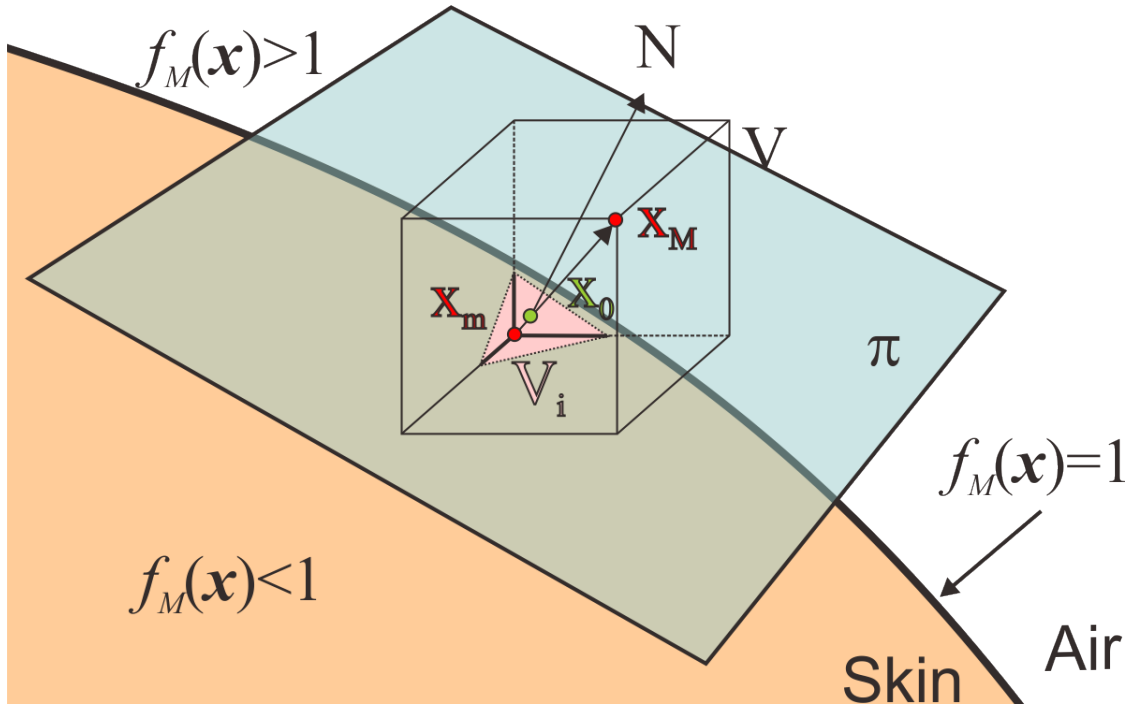
Figure 26: Partial volume approximation on skin/air boundary

For voxels on ligament/compartment boundary (where the compartment can correspond to glandular or adipose tissue), we approximated the ligament/compartment boundary using the gradients,, see the figure below.
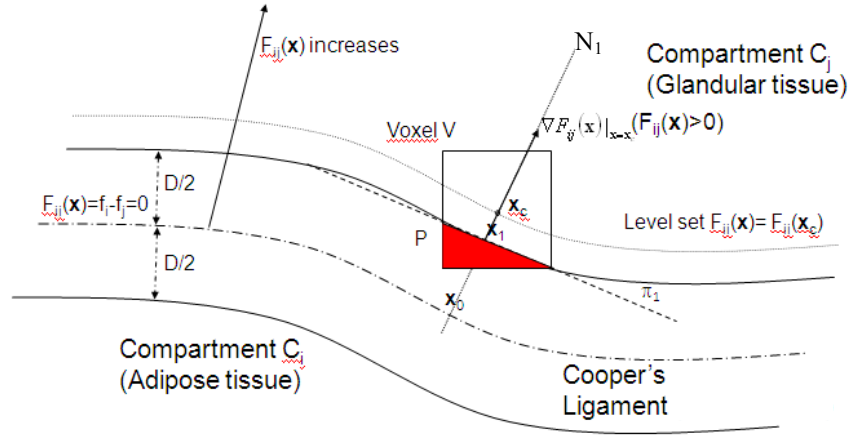
Figure 27: Planar approximation of a boundary between Cooper's ligament and a compartment (IEEE TMI, submitted)

Specifically, the approximation is defined as

$$\pi_1 : \left(\mathbf{x} - \mathbf{x}_1\right) \cdot \mathbf{N}_1 = 0 .$$

Where:

$$\mathbf{x}_1 = \mathbf{x}_c + sign\left(F_{ij}\left(\mathbf{x}_c\right)\right)\left(D/2 - \frac{F_{ij}\left(\mathbf{x}_c\right)}{\left\|\nabla F_{ij}\left(\mathbf{x}_c\right)\right\|}\right) \cdot \frac{\nabla F_{ij}\left(\mathbf{x}_c\right)}{\left\|\nabla F_{ij}\left(\mathbf{x}_c\right)\right\|},$$

$$\mathbf{N}_1 = sign\left(F_{ij}\left(\mathbf{x}_c\right)\right)\nabla F_{ij}\left(\mathbf{x}_c\right).$$

$\mathbf{x}_c$ is a voxel center, $F_{ij}(\mathbf{x})=0$ is a median surface, $\mathbf{N}_1$ is a vector normal to a level set $F_{ij}(\mathbf{x})= F_{ij}(\mathbf{x}_c)$ and $\mathbf{x}_1$ is a point on the normal at distance D/2 from the intersection of the normal and the median surface (D is a nominal thickness of a simulated ligament).

For voxels containing two materials, the partial volume computation subsequently reduces to computing a volume of a cube below (or above) a plane. We considered all possible cases for plane/cube alignment (see the next figure).
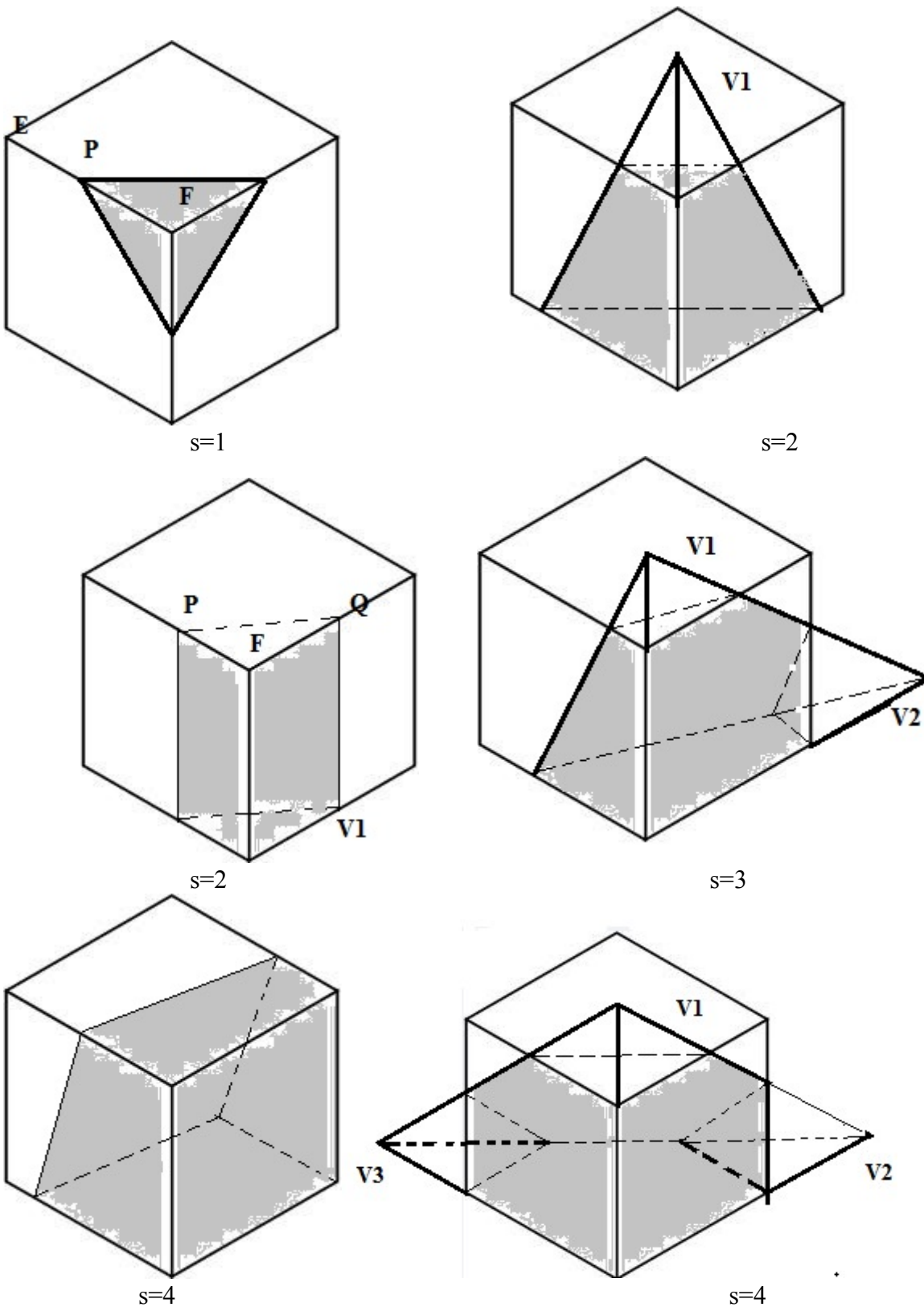
Figure 28: Computing partial volume of a cube below/above a plane using geometric primitives

Then, depending on the number of cube vertices above a planar approximation, we executed the algorithm drafted below.

$Vi = PV\_compute\_2(\mathbf{Q}_{i,i=1,...nVertex}, \Delta\mathbf{x}, \mathbf{N}, \mathbf{x}_0, nVertex)$

// **Inputs**: Voxel vertices $\mathbf{Q}_i, i=1,\ldots,nVertex$ above plane π ($nVertex \leq 4!$)

// voxel linear dimension Δx
// normal $\mathbf{N}$ and a point $\mathbf{x}_0$ specifying plane π.
// **Output:** Partial volume Vi of the voxel above the plane π
IF   nVertex==0
        RETURN 0.   //Partial volume is 0
ELSEIF nVertex==1   //Volume is a right angle triangular pyramid. CASE A (see Fig. 28, s=1)
        COMPUTE *nPoint=2*nVertex+1.* //the number of intersections.
        COMPUTE intersections $\mathbf{P, Q, R}$ between π and e_i, the edges containing $\mathbf{Q}_1$.
        COMPUTE the distances between intersections and $\mathbf{Q}_1$.
        RETURN volume of a tetrahedron $\mathbf{PQRQ}_1$. //(the shaded tetrahedron in Fig. 28, s=1)
ELSEIF nVertex==2
        DETERMINE the edge e containing $\mathbf{Q}_1$ and $\mathbf{Q}_2$
        IF $\mathbf{N} \perp e$ //Volume is a triangular prism. CASE B1
            COMPUTE intersections $\mathbf{P}$, $\mathbf{Q}$ between π and edges containing $\mathbf{Q}_1$ (except e).
            RETURN volume of a prism defined with base $\mathbf{PQQ}_1$ and height e.
            //(the shaded prism in Fig. 28, s=2)
        ELSE   //Volume is a triangular cut pyramid. CASE B2
            COMPUTE *nPoint=2*nVertex+1.*   //the number of intersections.
            COMPUTE intersections $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{S}$, $\mathbf{T}$ between π and each edge (or extension of edge) containing
            $\mathbf{Q}_1$ or $\mathbf{Q}_2$.
            RETURN volume difference b/w tetrahedra $\mathbf{PQRQ_1}$, $\mathbf{PSTQ_2}$. // (the shaded non-prism solid part in
Fig. 28, s=2)
        ENDIF
ELSEIF nVertex==3   //Volume is a "double cut" pyramid. CASE C
        COMPUTE *nPoint=2*nVertex*+1.   // the number of intersections.
        COMPUTE intersections $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{S}$, $\mathbf{T}$, $\mathbf{U}$, $\mathbf{W}$ between π and each edge (or extension of edge)
        containing $\mathbf{Q}_1$, $\mathbf{Q}_2$ or $\mathbf{Q}_3$.
            RETURN volume difference b/w tetrahedra $\mathbf{PQRQ_1}$, $\mathbf{PSTQ_2}$, $\mathbf{QUWQ_3}$. //(the shaded solid part in
Fig. 28, s=3)

ELSEIF nVertex==4
    IF vertices $\mathbf{Q}_i$ $i=1,\ldots,4$ are *coplanar* //Volume is a prismoid. CASE D1
            COMPUTE intersections $\mathbf{P,Q,R,S}$ between π and edges vertical to plane defined by $\mathbf{Q}_i$ $i=1,\ldots,4$.
            RETURN volume of prismoid $\mathbf{Q_1Q_2Q_3Q_4PQRS}$. //(the shaded prismoid in Fig. 28, s=4)

        ELSE   //Volume is a "triple cut" pyramid. CASE D2
            COMPUTE *nPoint=2*nVertex*+1. //The number of intersections.
            COMPUTE intersections $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{S}$, $\mathbf{T}$, $\mathbf{U}$, $\mathbf{W}$, $\mathbf{Y}$, $\mathbf{Z}$ between π and each edge (or extension of edge)
            containing $\mathbf{Q}_1$, $\mathbf{Q}_2$, $\mathbf{Q}_3$ or $\mathbf{Q}_4$.
            RETURN volume difference b/w tetrahedron $\mathbf{PQRQ}_2$ and tetrahedra $\mathbf{PSTQ_1}$, $\mathbf{QUWQ_3}$, and $\mathbf{RYZQ_4}$.
//(the shaded non-prismoid part in Fig. 28, s=4)

        ENDIF
ENDIF

    Algorithm 1: Algorithm for computation of partial volume of a voxel above a plane for different number of
                    vertices above the plane (IEEE TMI, submitted)

Note that this algorithm reduces the computation of partial volume to computation volume of a few primitives
(such as prisms and tetrahedrons) that is computationally efficient.

For voxels containing three materials, we utilized a method based on the application of the Gauss-Ostrogradsky theorem that can easily be generalized to a multi-material case. Assuming the partial volume V_i contains a voxel vertex Q_1, the volume is computed as:

$$\left| V_i \right| = \frac{(S_1 + S_2 + S_3)\Delta x + A_{\pi 1}d_1 + A_{\pi 2}d_2}{3}$$

where Δx is the length of the voxel side, S_1, S_2 and S_3 are surface areas of the boundaries of V_i belonging to the sides of the voxel that do not contain Q_1 and $A_{\pi 1}$, $A_{\pi 2}$ are surface areas of the parts of the planar approximations that bound V_i. see Figure below.
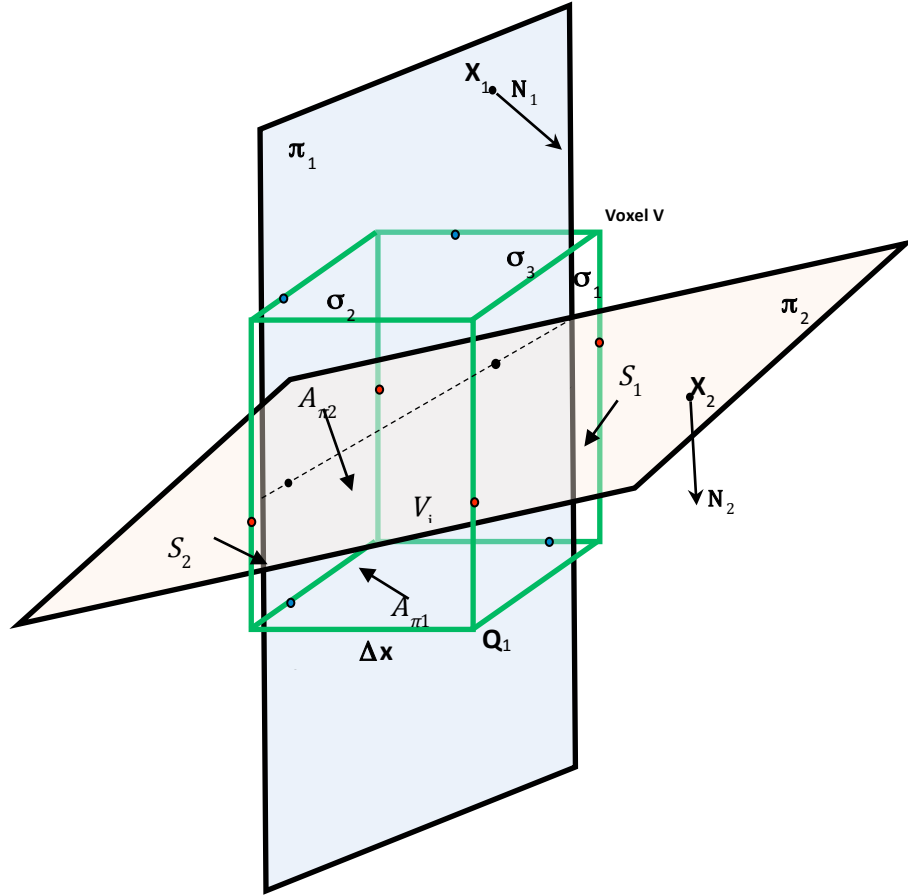
Figure 29: Computation of partial volume of voxels containing three materials using Gauss-Ostrogradsky theorem

We developed and implemented a Monte-Carlo based method for validation of partial volume computation. We demonstrated that the application of partial volume effects can lead to reducing artifacts in simulated mammograms, see Figure below.

(a)   (b)
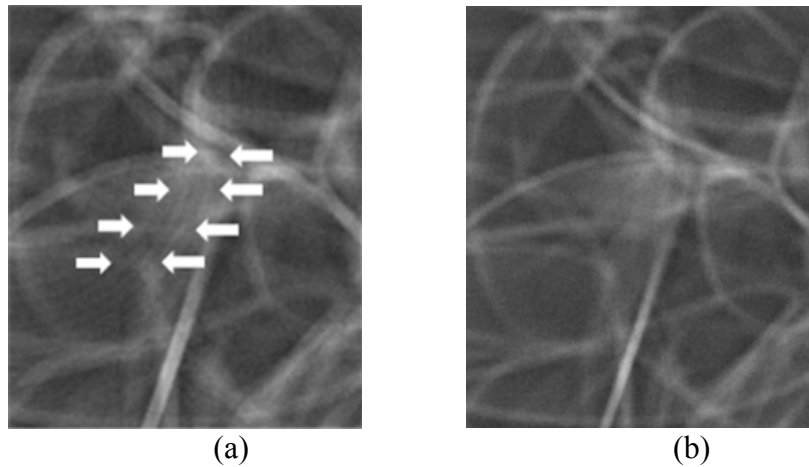
Figure 30: (a) Detail of a simulated projection of a 200um voxel phantom with no partial volume simulated; arrows indicate the artifacts; (b) Detail of the equivalent projection of a partial volume phantom (IEEE TMI, submitted).

Software Pipeline for Breast Anatomy and Imaging Simulation

We designed software pipeline that integrates breast simulation. The pipeline connects anatomy and imaging simulation components, necessary for the performance of virtual clinical trials, for preclinical validation of breast imaging systems. The components are connected using the XML-based parsimonious data representation. Optimized simulation algorithms and their GPU-based implementation allow very fast (practically real time) simulation, supporting virtual trials with very large number of simulated anatomies. The pipeline design and simulation were discussed in our publications at 2013 AAPM, 2013 RSNA, and 2014 SPIE conferences.
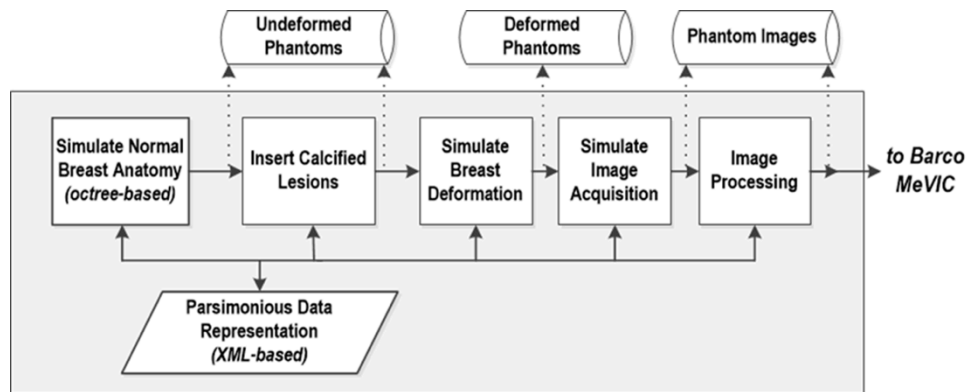
Figure 31: Flowchart of the Software Pipeline for Breast Anatomy and Imaging Simulation

o   **Prepare peer-review publications on the results of the proposed research. (Y3-Y4)**

While working on the current research, we have prepared several publications about our results. These publications are listed in the section on "Reportable Outcomes".

o **Validate success of the research training program by annual teleconferences with and bi-annual visits by external Advisory Committee.**

DSU faculty met with UPENN mentors on March 23, 2012, October 19, June 24, and January 24, 2011 to discuss the progress and the future work of each DSU faculty. Among additional collaborative activities between the Penn and DSU, Dr. Pokrajac took his sabbatical leave from DSU to work with Dr. Maidment and Dr. Bakic in Fall semester of 2011. Their collaborative work has generated several journal papers and conference publications (see Reportable Outcomes). They are also preparing to submit an NIH R01 grant proposal in Spring 2012 to the RFA on the Continued Development of Biomedical Software (**PAR-11-028**): http://grants.nih.gov/grants/guide/pa-files/PAR-11-028.html).

Our manuscript about the simulation of partial volume is conditionally accepted. Also, our manuscript on computational complexity of the recursive partitioning algorithm is accepted and published. In addition, two papers were accepted for the 2014 International Workshop on Breast Imaging (IWDM), held in Gifu, Japan in July 2014. The papers describe our preliminary results on the simulation of small scale tissue structures by subdivision of phantom adipose compartments, as well as the preliminary results in the correlation between topological and textural properties extracted from clinical breast images. Both properties have been previously related with the breast cancer risk, thus their correlation may offer an improved estimation of risk.

o **Validate success of the research training program by annual teleconferences with and bi-annual visits by external Advisory Committee.**

On November 3, 2010, a teleconference meeting of the DoD award Advisory Committee was organized by Drs. Maidment, Liu, and Bakic, and attended by all the DSU faculty supported on the grant, as well as Drs. Chanita Hughes and Timothy Rebbeck from UPenn. The discussed issues include our progress on the grant, future research steps related to the genetic analysis project aims, as well as the long-term aim of establishing a regional Breast Cancer Disparity Center at DSU.

DSU faculty met with UPENN mentors on August 9, 2010, January 24, 2011, June 24, 2011, and March 23, 2012 to discuss the progress and the future work of each DSU faculty.

2.3 Objective 3

Objective 3 (from grant SOW): Produce a corpus of high-quality published work and develop a portfolio of independently funded research grants at DSU to support a sustained breast cancer program

Our results from the proposed research and 34 published journal and conference papers (listed in Chapter 4) have been used as the basis for preparing and submitting the following funding applications:

- **June 2012, NIH R01 proposal** "Continued Development and Maintenance of Computer Simulation of Breast Anatomy," (D. Pokrajac, P. Bakic, A. Maidment) submitted to PAR-11-028:
  http://grants.nih.gov/grants/guide/pa-files/PAR-11-028.html;
  The application was scored at 41% but not funded.
- **July 2013, a resubmission of the NIH R01 proposal** "Continued Development and Maintenance of Breast Anatomy and Imaging Simulation," (D. Pokrajac, P. Bakic, A. Maidment):

The application was scored at 40% but not funded.
- **November 2013, DoD Breakthrough Award proposal** "Identification of Aggressive DCIS Cases by 3D Analysis of Reconstructed Digital Breast Tomosynthesis Images," (P. Bakic, D. Pokrajac);
The proposal was not funded.
- **October 2013, Delaware INBRE pilot proposal** "Improving Realism of Software Breast Phantoms" (D. Pokrajac);
The proposal has been funded in the amount of $160,000 for the duration of 2014-2016.

3. Important Findings

- We finalized ACRIN-DMIST data transfer on MIRC and resolved issues with various batch transfers including 11,106 anonymized, previously acquired cases;
- We developed a program to estimate the breast cancer risks using the metadata for women from the ACRIN database and the software implementation of the Gail risk model;
- We estimated breast densities for mammograms from the ACRIN database, using the software developed at the University of Pennsylvania;
- We developed a web-based data center for this project, which includes database design and merge of source data, converted data and computed data together into relational database tables
- We confirmed that the validity of the conjecture that breast density significantly depends on race
- We performed preliminary testing on improvement of breast density using the image based risk descriptors

- We performed a preliminary query of the ACRIN data aimed at identifying the prevalence of women with incomplete visualization of the breast;
- For the cases with only partial breast visualization, we have developed an image registration/fusion method for the estimation of biomarkers;
- We developed an efficient method for generating anthropomorphic software breast phantoms with high resolution;
- We proved computational complexity of the simulation algorithm and mathematically demonstrated its asymptotic efficiency
- Updated the design of the software pipeline for real-time simulation of breast anatomy and imaging.
- We developed a computer demo of the real-time simulation of breast anatomy and imaging
- We developed and validated several modifications and novel features to breast anatomy simulation methods, used for generating software breast phantoms, including: (1) a method to improve thickness control of the Cooper's ligaments in the simulation algorithm by reducing "dents" on the ligaments' surface; (2) a method for insertion of simulated microcalcification clusters in the software breast phantom; (3) a method for simulating the dense tissue distribution in the software phantom; (4) a method to simulate small scale tissue structure; and (5) a method for reducing quantization artifacts in phantom images by simulating partial volume in voxels containing several simulated tissue types. The listed novel features have resulted in further improvement of the image quality in synthetic breast images generated using the software phantom.
- We performed preliminary analysis of the correlation between topological and textural properties extracted from clinical breast images
- We developed preliminary algorithms for (1) modeling of ducts and (2) simulation of subcutaneous tissue.
- We developed a fully-automated software pipeline to perform quantitative analysis of breast tissue composition from multimodality digital breast images.
- A roadmap has been proposed to optimize the simulation of breast anatomy by parallel implementation, in order to reduce the time needed to generate software breast phantoms;

- As a first step towards the shape analysis of anatomical structures in the breast, we have analyzed an anthropomorphic software breast phantom. The shape analysis was performed by fitting ellipsoids to the simulated tissue compartments.

4. Reportable Outcomes

- Xia, S., Liu, F., Maidment, A.D.A., Bakic, P.R.: "Refinements to the Deformation Model of An Anthropomorphic Computer Generated Breast Phantom," Medical Physics 37: 3131, 2010. (Presented as a poster at the 2010 Annual Meeting of the American Association of Physicists in Medicine, Philadelphia, PA.)

- Bakic, P.R., West, E., Sak, M., Gavenonis, S.C., Duric, N., Maidment, A.D.A.: "Comparison of 3D and 2D Breast Density Estimation from Synthetic Ultrasound Tomography Images and Digital Mammograms of Anthropomorphic Software Breast Phantoms," SPIE Medical Imaging 2011.

- Liu, F., Bakic, P.R., Pokrajac, D., Wilson, C., Shi, X., and Maidment, A.D.A: "Image-Based Biomarkers of Breast Cancer Risk: Analysis of Risk Disparity among Minority Populations", Era of Hope Breast Cancer Conference, Orlando, Fl, August 2-5, 2011.

- Pokrajac, D., Maidment, A.D.A., Bakic, P.R.: "A method for Fast Generation of High Resolution Software Breast Phantom," submitted for presentation at the 2011 Joint Annual Meeting of the American Association of Physicists in Medicine and Canadian Organization of Medical Physicists, Vancouver, BC.

- Chen, F., Pokrajac, D.D., Shi, X., Liu, F., Maidment, A.D.A., Bakic, P.R., "Partial Volume Simulation in Software Breast Phantoms," Proc. SPIE 8313, Medical Imaging 2012: Physics of Medical Imaging, 83134U, 2012.

- Chui, J.H., Pokrajac, D.D., Maidment, A.D.A., Bakic, P.R., "Roadmap for Efficient Parallelization of Breast Anatomy Simulation," Proc. SPIE. 8313, Medical Imaging 2012: Physics of Medical Imaging, 83134T, 2012.

- Contijoch, F., Lynch, J., Pokrajac, D.D., Maidment, A.D.A., Bakic, P.R., "Shape Analysis of Simulated Breast Anatomical Structures," Proc. SPIE. 8313, Medical Imaging 2012: Physics of Medical Imaging, 83134J, 2012.

- Chen, F., Pokrajac, D.D., Shi, X., Liu, F., Maidment, A.D.A., Bakic, P.R., "Simulation of Three Material Partial Volume Averaging in Software Breast Phantom," In Maidment A., Bakic P., and Gavenonis S. (eds.), Breast Imaging (IWDM), Lecture Notes in Computer Science 7361, 149-156, Springer-Verlag Berlin Heidelberg, 2012.

- Chui, J.H., Pokrajac, D.D., Maidment, A.D.A., Bakic, P.R., "Toward Breast Anatomy Simulation Using GPU," In Maidment A., Bakic P., and Gavenonis S. (eds.), Breast Imaging (IWDM), Lecture Notes in Computer Science 7361, 506-513, Springer-Verlag Berlin Heidelberg, 2012.

- Pokrajac, D., Maidment, A.D.A., Bakic, P.R., "Optimized Generation of High Resolution Breast Anthropomorphic Software Phantoms," Medical Physics 39, 2290, 2012.

- Bakic, P.R., Maidment, A.D.A., Chui, J.H., Avanaki, A.N., Marchessoux, C., Pokrajac, D.D, Espig, K., Kimpe, T., Xthona, A., Lago, M., Shankla, V.: "Automated and optimized imaging simulation platform for virtual clinical trials of breast cancer screening." In Proc. of the Scientific Assembly and Annual Meeting of the Radiological Society of North America, 2013.

- Bakic, P., Myers, K., Reiser, I., Kiarashi, N., Zeng., R.: "Virtual Tools for Validation of X-ray Breast Imaging Systems," Medical Physics, 40: 3133. 2013.

- Chui, J.H., Zeng, R., Pokrajac, D.D., Park, S., Myers, K.J., Maidment A.D.A., Bakic, P.R., "Two Methods for Simulation of Dense Tissue Distribution in Software Breast Phantoms," Proc. SPIE. 8668, Medical Imaging 2013: Physics of Medical Imaging, 86680M, 2013.

- Chen, F., Zheng, P., Xu, P., Maidment, A.D.A., Bakic, P.R., Pokrajac, D.D., Liu, F., Shi, X., "Breast image registration by using non-linear local affine transformation," Proc. SPIE 8668, Medical Imaging 2013: Physics of Medical Imaging, 86684J, 2013.

- Maidment, A.D.A., Bakic, P.R., Chui, J.H., Avanaki, A.N., Marchessoux, C., Pokrajac, D.D, Espig, K., Kimpe, T., Xthona, A., Lago, M., Shankla, V.: "The role of virtual clinical trials in preclinical testing of breast imaging systems." In Proc. of the Scientific Assembly and Annual Meeting of the Radiological Society of North America, 2013.

- Abbey, C.K., Bakic, P.R., Pokrajac, D.D., Maidment, A.D.A., Eckstein, M.P., Boone, J.M., "Non-Gaussian Statistical Properties of Virtual Breast Phantoms," In Image Perception, Observer Performance, and Technology Assessment, Proc. SPIE 9037, ed. by C. Mello-Thoms, M. Kupinski, 90370G, 2014.

- Bakic, P.R., Pokrajac, D.D., De Caro, R., Maidment, A.D.A, "Realistic Simulation of Breast Tissue Microstructure in Software Anthropomorphic Phantoms," In Fujita H., Hara T., Muramatsu, C. (eds.), Breast Imaging (IWDM), Lecture Notes in Computer Science 8539, 348-355, Springer-Verlag Berlin Heidelberg, 2014.

- Bakic, P.R., Pokrajac, D.D., Thomas, M., Skoura, A., Nuzhnaya, T., Megalooikonomou, V., Keller, B., Zheng, Y., Kontos, D., Gee, J.C., Cardenosa, G., Maidment, A.D.A., "Correlation Between Topological Descriptors of the Breast Ductal Network from Clinical Galactograms and Texture Features of Corresponding Mammograms," In Fujita H., Hara T., Muramatsu, C. (eds.), Breast Imaging (IWDM), Lecture Notes in Computer Science 8539, 658-665, Springer-Verlag Berlin Heidelberg, 2014.

- Shankla, V., Pokrajac, D.D., Weinstein, S.P., Conant, E.F., Maidment, A.D.A., Bakic, P.R.: "Automatic insertion of simulated microcalcification clusters in a software breast phantom," In Physics of Medical Imaging, Proc. SPIE 9033, ed. by B. Whiting, C. Hoeschen, D. Kontos, 2014.

- T. Nuzhnaya, A. Skoura, G. Cardenosa, V. Megalooikonomou, D. Kontos, D. Pokrajac, P. Bakic, A. Maidment, "Correlation between topological descriptors of the breast ductal network extracted from clinical galactograms and texture features of corresponding mammograms," *Proc. IWDM 2014*.

- Petkovic, M., Bakic, P.R., Maidment, A.D.A., Pokrajac, D.D., "Asymptotic Number of Z3Δ Cells Covering C(1) Surface on Uniform Grid and Complexity of Recursive-Partitioning Simulation of Septal Tissue Regions," Applied Mathematics and Computation 252-263, 2015.

- Chen, F., Bakic, P.R., Maidment, A.D.A., Jensen, S.T., Shi, X., Pokrajac, D.D., "Description and Characterization of a Novel Method for Partial Volume Simulation in Software Breast Phantoms," conditionally accepted for publication in IEEE Transactions on Medical Imaging, 2015.

Presentations:

- Recursive Partitioning for Simulation of Breast Tissue, Faculty of Sciences, University of Nis, Serbia, 10/6/2011 (Pokrajac, Bakic)

- Novel Algorithm for Breast Anatomy Simulation, Optimized for Generation of High Resolution Software Phantoms, MIPG Seminar Series, University of Pennsylvania, 10/21/2011 (Pokrajac, Maidment, Bakic)

- Simulation of Breast Tissue using Recursive Partitioning Algorithm, Faculty of Electrical Engineering, University of Nis, Serbia, 12/27/2011 (Pokrajac, Bakic)

- Breast Tissue Simulation with Recursive Partitioning Algorithm – Latest results, University of Belgrade, Serbia, 3/5/2012 (Pokrajac, Bakic)

- "Mathematics in the Chest", Faculty of Electrical Engineering, Nis, Serbia, 3/30/2012 (Pokrajac; Sponsored by American Embassy in Serbia)

- Partial volume simulation in software breast phantoms, Mid-Atlantic Numerical Analysis Day, Temple University, Philadelphia, PA, November 4, 2011 (Chen, Pokrajac, Shi, Liu, Maidment, Bakic)

- D. Pokrajac made a presentation at the IWDM 2012, the 11th International Workshop on Breast Imaging, in Philadelphia, PA. on July 8-11, 2012, "Toward Breast Anatomy Simulation using GPU" (co-authored by J. Chu, D. Pokrajac, A. D. Maidment, P. Bakic)
- F. Chen made a presentation at the IWDM 2012, the 11th International Workshop on Breast Imaging, in Philadelphia, PA. On July 8-11, 2012, "Simulation of Three Materials Partial Volume Averaging in a Software Breast Phantom" (co-authored by F. Chen, D. Pokrajac, X. Shi, F. Liu, A. D. Maidment, P. Bakic)

- Breast Tissue Simulation with Recursive Partitioning Algorithm, Faculty of Electronics, Mechanical Engineering and Ship building, Split, Croatia, May 21, 2014 (Pokrajac, Bakic, Maidment), invited talk

- Simulation of Breast Tissues using Computer Algorithms, Kolarac Institution, Belgrade, Serbia, May 30. 2014 (Pokrajac, Bakic, Maidment), invited talk

- Partial volume simulation in software breast phantoms, XIII Serbian Mathematical Congress, Vrnjacka Banja, Serbia, May 23, 2014 (Pokrajac, Bakic, Maidment, Shi, Chen)

- Mathematical Issues in Software Breast Phantom Simulation, XIII Serbian Mathematical Congress, Vrnjacka Banja, Serbia, May 23, 2014 (Pokrajac, Maidment, Petkovic, Bakic), invited talk

5. Conclusion

Supported by this DoD HBCU Partnership Training Award, our team of DSU and UPENN researchers have designed and successfully completing a breast cancer research training program for DSU faculty, and performed a broad set of research activities focused on the analysis of breast cancer imaging, cancer risk estimation from demographic and imaging based descriptors, mammographic image processing and registration, and simulation of breast tissue anatomy and imaging. We have organized the well-attended bi-weekly DSU-UPENN Breast Cancer Seminar series during the Y1 and Y2. Our research results have been presented in 34 conference and

journal publications. DSU's rank was moved to number 13 of all HBCUs partly due to the recent research activities including this funded project by DoD. A total of five graduate and undergraduate students at DSU and UPENN have worked on this project; one Ph.D. dissertation (Chen) has been successfully defended. Finally, in 2014, our grant application to NIH-supported Delaware INBRE program was funded. This award has helped significantly to establish the breast cancer research program at DSU, as well as to successfully apply for future research funding. Through this recently approved funding, the DSU-UPENN research partnership, initiated through our DoD HBCU PTA award, will continue in the future!

6. Appendices (New publications since the last annual report)

- Bakic, P.R., Pokrajac, D.D., Thomas, M., Skoura, A., Nuzhnaya, T., Megalooikonomou, V., Keller, B., Zheng, Y., Kontos, D., Gee, J.C., Cardenosa, G., Maidment, A.D.A., "Correlation Between Topological Descriptors of the Breast Ductal Network from Clinical Galactograms and Texture Features of Corresponding Mammograms," In Fujita H., Hara T., Muramatsu, C. (eds.), Breast Imaging (IWDM), Lecture Notes in Computer Science 8539, 658-665, Springer-Verlag Berlin Heidelberg, 2014;

- Petkovic, M., Bakic, P.R., Maidment, A.D.A., Pokrajac, D.D., "Asymptotic Number of Z3Δ Cells Covering C(1) Surface on Uniform Grid and Complexity of Recursive-Partitioning Simulation of Septal Tissue Regions," Applied Mathematics and Computation 252-263, 2015;

- Chen, F., Bakic, P.R., Maidment, A.D.A., Jensen, S.T., Shi, X., Pokrajac, D.D., "Description and Characterization of a Novel Method for Partial Volume Simulation in Software Breast Phantoms," accepted for publication in IEEE Transactions on Medical Imaging, 2015.

Correlation between Topological Descriptors of the Breast Ductal Network from Clinical Galactograms and Texture Features of Corresponding Mammograms

Predrag R. Bakic[1], David D. Pokrajac[2], Mathew Thomas[1], Angeliki Skoura[3], Tatyana Nuzhnaya[4], Vasileios Megalooikonomou[3,4], Brad Keller[1], Yuanjie Zheng[1], Despina Kontos[1], James C. Gee[1], Gilda Cardenosa[5], and Andrew D.A. Maidment[1]

[1] Dept. of Radiology, University of Pennsylvania, Philadelphia, PA, USA
[2] Computer and Information Sciences Dept., Delaware State University, Dover, DE, USA
[3] Computer Engineering and Informatics Dept., University of Patras, Greece
[4] Computer and Information Science Dept., Temple University, Philadelphia, PA, USA
[5] Dept. of Radiology, Virginia Commonwealth University, Richmond, VA, USA
Predrag.Bakic@uphs.upenn.edu

Abstract. Mammographic texture has been reported as a biomarker of cancer risk. Recent publications also suggest correlation between the topology of the breast ductal network and risk of cancer. The ductal network can be visualized by galactography, the preferred imaging technique for nipple discharge. We present current results about the correlation between topological and textural properties of clinical breast images. This correlation was assessed for 41 galactograms and 56 mammograms from 13 patients. Topology was characterized using feature extraction techniques arising from text-mining, validated previously in the classification of normal, benign, and malignant galactograms. In addition, we calculated 26 texture descriptors using an automated breast image analysis pipeline. Regression analysis was performed between texture and topological descriptors averaged over all images of the same patient. These data demonstrate a correlation between topology and a subset of texture features with borderline statistical significance due to the limited sample size.

Keywords: Texture analysis, topology descriptors, galactograms, mammograms.

1   Introduction

Previously, we analysed the topological properties of the branching network of breast ducts as visualized by galactography, an x-ray imaging procedure of the contrast-enhanced breast ductal network (1-3). That analysis suggested a correlation between cancer risk and ductal network topology; this correlation also has been supported by evidence from murine cancer models (4). Clinical visualization of breast ducts is, however, not routinely performed; galactography is indicated infrequently, and it mostly commonly reveals benign findings (5, 6).

On the other hand, texture descriptors of breast parenchyma are known to correlate with cancer risk (7-9). Our work is motivated by a desire to determine whether there is an association between parenchymal texture descriptors and ductal topology. Such an analysis would lead to improved models of breast anatomy, and may lead to a better understanding of breast cancer risk. Currently, breast cancer risk is estimated using patient demographic information and parenchymal texture features extracted from 2D mammograms. The spatial arrangement of breast tissue is, however, three-dimensional, stressing the need to understand the relationship between parenchymal structure and image texture.

The UPenn X-ray Physics Lab has extensive experience with the simulation of breast anatomy and imaging (10, 11). The development of the UPenn breast phantom is predicated upon a set of anatomically justified elements. To that end, we have chosen not to model the parenchymal texture by a random field with statistical properties similar to clinical data. This development process has been incremental, and continues to this day. For example, we have just recently begun to model the hierarchical organization of Cooper's ligaments seen in breast histology slices. A preliminary validation of a model of this small scale tissue detail, published separately in this proceedings, indicates good agreement with clinically estimated texture (12).

This paper presents our current results about the correlation between the ductal topology of clinical galactograms and the parenchymal textural properties of clinical mammograms from the same group of women. Understanding the relationship between mammographic texture and spatial distribution of breast anatomy will help optimize and extend our fully automated software pipeline for breast anatomy and imaging simulation; ultimately, we would like to be able to simulate specific cohorts of women, stratified by age, risk, and other factors.

2   Methods

2.1   Topological Analysis of Galactograms

In this paper, we analysed images of existing, anonymized clinical galactograms of 49 women, obtained from Virginia Commonwealth University. The data collection was performed after IRB review and was HIPAA compliant. Clinical galactograms were digitized from film, and categorized based upon the visibility of the ductal network. Ductal trees were traced manually from galactograms, followed by Prufer encoding of the breadth-first labelled ductal tree nodes (3). Then tf-idf significance weighting (3), originally used in text mining, was performed on the traced and encoded ductal trees. After manually tracing the ductal networks, a subset of 41 galactograms from 13 patients with well-defined ductal trees was selected for further processing and testing.

2.2   Texture Analysis of Mammograms

We measured 26 texture features in 56 digitized mammograms from the 13 selected patients imaged at Virginia Commonwealth University. Texture analysis was performed

using a fully automated software pipeline which extracted a large set of image features from the digitized mammograms (13). The pipeline calculates texture feature maps at points on a regular spatial lattice, determined by two parameters: the window size and the lattice distance. Here we use a window size of 63 pixels, and a lattice distance of 31 pixels. The analysed features are organized into three groups, including (i) descriptors of grey-level histograms, (ii) co-occurrence features, and (iii) run length features. These texture features have been used previously in breast cancer risk assessment studies (9). For the correlation analysis, the texture feature maps were averaged over the whole breast region (excluding the pectoral muscle and air).

2.3   Hypothesis Testing

We tested the hypothesis that there is a correlation between mammographic texture features and ductal topology descriptors. To that end, we have calculated the linear regression (14). The goal was to predict values of texture features averaged over all mammograms of the same patient as a function of the topological properties estimated from the corresponding manually-traced ductal networks, averaged over all galactograms of the same patient. Prior to the regression analysis, we combined the tf-idf topological descriptors via principal component analysis (PCA). The regression model considered the first 13 PCA components and the 26 texture features.

2.4   Power Calculations

It can be demonstrated that a small sample size (in this case 13 patients), could lead to large estimated p-values and hence rejection of valid linear regression models (large Type II error). To demonstrate the effect of sample size, we simulated an augmented dataset by bootstrapping (15). The bootstrapping was performed by replicating data records, with added Gaussian noise, for each PCA attribute and response variable. The standard deviation of the noise was 50% of the estimated standard deviation of the attributes or response variables.

3   Results and Discussion

Fig. 1 shows an example of a clinical galactogram used in this study (Fig. 1(a)), and the corresponding manually-traced ductal tree (Fig. 1(b)). The Prufer encoding and the tf-idf weights corresponding to the traced tree is also given (Fig. 1(c-d)). The example shown illustrates a breast with a malignant finding. Fig. 2 shows an example of a clinical mammogram from the same woman (Fig. 2(a)) and the corresponding texture feature map (Fig. 2(b)). Shown in this example, is a map of the entropy texture feature.
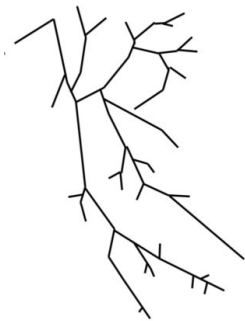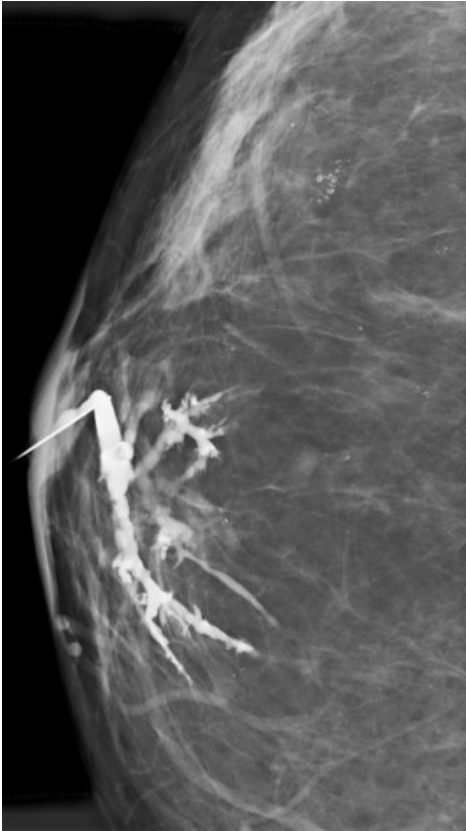
(a)   (b)

```
[ 1   1   3   4   6  10  10   6  11  16  16  11  17  24  24  17  25  34
 34  47  50  50  47  25   4   7  12  18  26  26  18  27  38  38  49  49
 27  12   7  13  20  28  28  20  29  29  13  21  30  30  21   3   5   5]
```

(c)

```
[0.00 0.11 0.45 0.36 0.40 0.50 0.71 0.00 0.50 0.77
 1.04 0.77 0.97 1.11 0.83 0.00 0.00 1.04 1.55 1.27
 1.45 1.19 0.00 0.00 1.27 0.00 0.00 0.00 0.00 0.00
 0.00 1.27 0.00 0.00 1.66 1.55 0.00 0.00 1.55 0.00
 0.00 1.45 0.00 2.04 2.19 1.66 0.00 1.90 0.00 0.00
 2.19 0.00 0.00 0.00 0.00 1.55 0.00 0.00 0.00 3.36]
```

(d)

Fig. 1. Illustration of the topological descriptors of the breast ductal network. Shown are: (a) a clinical galactogram with malignant finding; (b) the corresponding manually-traced ductal tree; (c) the Prufer encoding; and (d) the tf-idf weights corresponding to the ductal tree.
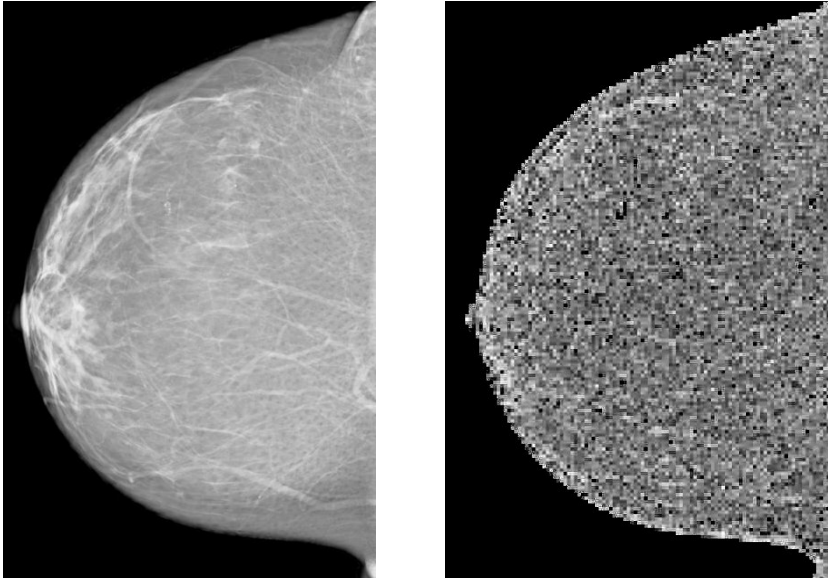
Fig. 2. The clinical mammogram of the patient from Fig. 1 (left) and the corresponding map of the entropy texture feature (right)

Fig. 3 shows the regression analysis results. A borderline statistical significance was observed for three texture features (features 9-11, p-values between 0.05 and 0.1); three additional features (features 17, 20, 24) had p-values between 0.1 and 0.15. It is likely that statistical significance was not achieved due to the limited sample size; thus, we are prevented from drawing a definitive conclusion about the correlation between texture and topology. The observed results, however, suggest a possible correlation between topological descriptors and several texture features. The bootstrap analysis of a hypothetically enlarged dataset with a sample size of 26 suggests that a statistically significant regression (at a significance level of 0.05) could be achieved between various texture features and topological descriptors. Fig. 3 shows the p-values from the bootstrap analysis.

The potential for inter-correlation between individual texture features was accounted for by applying PCA before performing the regression analysis, as PCA uses orthogonal transformations to convert the original data into a set of linearly uncorrelated variables. The bootstrap analysis performed in this paper to estimate the effect of sample size, assumed the noise in the enlarged data set to have a standard deviation equal to 50% of the standard deviation in individual sample data.

The results presented in this paper are based upon an initial analysis of 13 patients. We are currently analysing a larger set of clinical breast images; we expect to double the sample size in the near future. If the linear dependence between the texture and topology is confirmed (as suggested from our initial analysis and supported by bootstrapping), texture descriptors could be used as a proxy for topology, since the ductal network is not routinely visible in clinical images. Identifying texture features, or

combinations of texture features, which have the strongest correlation with topology could improve the understanding of texture-based risk biomarkers.

If, however, the increased sample size does not confirm the correlation between topology and texture, it could suggest that topology may carry risk-related information independent from texture descriptors. This could potentially lead to an improvement in the accuracy of breast cancer risk estimation techniques, assuming a clinically feasible method for the visualization and characterization of breast ducts (e.g., MRI or tomosynthesis) is available.
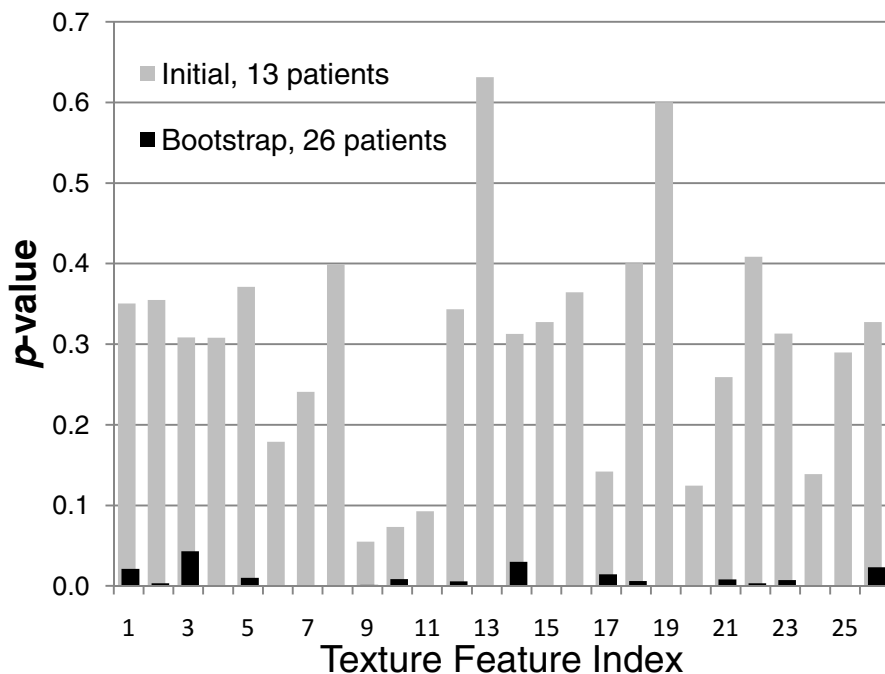
Fig. 3. *p*-value for the regression of individual texture features (averaged over all mammograms of the same patient) as a function of principal component analysis (PCA) components for the topological descriptors (tf-idf weights, averaged of all the traced ductal networks of the same patient). Shown are the results of the initial analysis of 13 patients, as well as the bootstrap results modelling a dataset of 26 cases.

It is worth noting the limitations of the current study. First, the ductal trees analysed in this paper were manually-traced from digitized galactograms. The manual tracing was performed by one person (a third-year medical student with experience in breast imaging). We believe that manual tracing did not compromise the analysis. In our previous study of ductal topology, we observed relatively low variations (a root-mean-square fractional error on the order of 2%) in estimated topological features due to manual tracing (2).

Additional potential limitations include the use of average texture descriptors, and the inter-correlation between individual descriptors of texture (or topology). In this paper, the regression analysis was performed using texture features averaged over the breast region in each mammographic image. These average values may suppress the differences in feature histograms calculated over mammographic images. In the future, we may repeat the analysis based upon other histogram moments, or using the full histogram as the texture descriptor.

4   Conclusion

We have performed a regression analysis between topological descriptors of the breast ductal network extracted from previously acquired, anonymized clinical galactograms, and texture descriptors estimated from corresponding clinical mammograms. Ductal networks were extracted from galactograms by manual tracing. The texture features were estimated using a fully automated image analysis pipeline. Initial analysis of clinical images from 13 women suggests correlation with borderline significance for a subset of texture descriptors. The identified subset of texture descriptors could hypothetically be used as proxy for ductal topological properties. Analysis of a larger number of clinical cases is ongoing.

References

[1] Bakic, P.R., Albert, M., Maidment, A.D.: Classification of galactograms with ramification matrices: Preliminary results. Academic Radiology 10(2), 198–204 (2003)
[2] Bakic, P.R., Albert, M., Brzakovic, D., Maidment, A.: Mammogram synthesis using a three-dimensional simulation. III. Modeling and evaluation of the breast ductal network. Medical Physics 30(7), 25–1914 (2003)
[3] Megalooikonomou, V., Barnathan, M., Kontos, D., Bakic, P.R., Maidment, A.D.A.: A Representation and Classification Scheme for Tree-Like Structures in Medical Images: Analyzing the Branching Pattern of Ductal Trees in X-ray Galactograms. IEEE Transactions on Medical Imaging 28(4), 93–487 (2009)
[4] Atwood, C.S., Hovey, R.C., Glover, J.P., Chepko, G., Ginsburg, E., Robison, W.G., et al.: Progesterone induces side-branching of the ductal epithelium in the mammary glands of peripubertal mice. Journal of Endocrinology 167(1), 39–52 (2000)

[5] Dinkel, H.P., Trusen, A., Gassel, A.M., Rominger, M., Lourens, S., Muller, T., et al.: Predictive value of galactographic patterns for benign and malignant neoplasms of the breast in patients with nipple discharge. British Journal of Radiology 73(871), 14–706 (2000)

[6] Cardenosa, G., Doudna, C., Eklund, G.: Ductography of the breast: technique and findings. America Journal of Roentgenology 162, 7–1081 (1994)

[7] Wolfe, J.N.: Breast patterns as an index of risk for developing breast cancer. American Journal of Roentgenology 126(6), 7–1130 (1976)

[8] Li, H., Giger, M.L., Huo, Z., Olopade, O.I., Lan, L., Weber, B.L., et al.: Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: Effect of ROI size and location. Medical Physics 31(3), 55–549 (2004)

[9] Kontos, D., Bakic, P.R., Carton, A.-K., Troxel, A.B., Conant, E.F., Maidment, A.D.A.: Parenchymal Pattern Analysis in Digital Breast Tomosynthesis: Towards Developing Imaging Biomarkers of Breast Cancer Risk. Academic Radiology 16(3), 98–283 (2008)

[10] Bakic, P.R., Albert, M., Brzakovic, D., Maidment, A.D.A.: Mammogram synthesis using a 3D simulation. I. Breast tissue model and image acquisition simulation. Medical Physics 29(9), 9–2131 (2002)

[11] Pokrajac, D.D., Maidment, A.D.A., Bakic, P.R.: Optimized generation of high resolution breast anthropomorphic software phantoms. Medical Physics 39(4), 302–2290 (2012)

[12] Abbey, C.K., Bakic, P.R., Pokrajac, D.D., Maidment, A.D.A., Eckstein, M.P., Boone, J.: Non-Gaussian Statistical Properties of Virtual Breast Phantoms. In: Mello-Thoms, C.R., Kupinski, M.A. (eds.) SPIE Image Processing, Observer Performance, and Technology Assessment, vol. 9037. SPIE, San Diego (2014)

[13] Zheng, Y., Wang, Y., Keller, B.M., Conant, E.F., Gee, J.C., Kontos, D.: A fully-automated software pipeline for integrating breast density and parenchymal texture analysis for digital mammograms: Parameter optimization in a case-control breast cancer risk assessment study. In: Novak, C.L., Aylward, S. (eds.) SPIE Computer-Aided Diagnosis, vol. 8670. SPIE, Lake Buena Vista (2013)

[14] Devore, J.: Probability and Statistics for Engineering and the Sciences. Brooks/Cole, Belmont (2008)

[15] Efron, B.: The jackknife, the bootstrap and other resampling plans, Philadelphia, PA (1982)

CrossMark
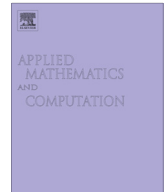
Asymptotic number of $\mathbb{Z}^3\Delta$ cells covering $\mathcal{C}^{(1)}$ surface on uniform grid and complexity of recursive-partitioning simulation of septal tissue regions

Marko D. Petković [a,*], Predrag R. Bakic [b], Andrew D.A. Maidment [b], David Pokrajac [c]

[a] University of Niš, Faculty of Science and Mathematics, Višegradska 33, 18000 Niš, Serbia
[b] Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA
[c] Computer and Information Sciences Department, Applied Mathematics Research Center, CREOSA Center, Delaware State University, Dover, DE 19901, USA

ARTICLE INFO

ABSTRACT

The exact asymptotic computational complexity for a problem of indexing cells on a uniform grid intersecting with a union of $\mathcal{C}^{(1)}$ surfaces has been proven. The computational complexity of the recursive partition indexing algorithm, utilized for simulation of septated tissues, is derived and the algorithm is demonstrated as being asymptotically optimal.

© 2014 Published by Elsevier Inc.

1. Introduction

Octrees (e.g., [1–3]), a 3D spatial indexing technique, have successfully been used in various applications in imaging and computer graphics. For example, octrees are utilized for efficient data representation in fast interactive rendering of isosurfaces [4], rendering of medical data [5,6], 3D surface-based thinning algorithms [7] and compression of complex isosurfaces [8]. Recent parallel applications of octrees include [9,10].

The need for preclinical validation and optimization of medical imaging systems or image analysis methods has recently led to the development of a recursive partitioning based simulation technique (e.g., [10–15]). In this technique, an organ of interest is specified using a system of scalar fields in 3D space. Various anatomic constituents of the organ (compartments, septal regions separating compartments, skin, etc.) are modeled by indexing voxels or groups of voxels using octrees. The recursive partitioning stops when the linear dimension of a cubic subdomain is equal to the prespecified size or when the cubic subdomain contains material of a single type.

The use of octree-based recursive partitioning in breasts simulation has led to a number of significant accomplishments. The GPU implementation allows for near real-time modeling of the breast; software phantoms are generated at a rate of 7 breasts/minute using a voxel resolution of 50 micrometers [10]. In addition, the method makes it possible to simulate the breast, in whole or in part, at the cellular level [11]. Use of the accelerated simulation at high spatial-resolution provides an avenue for realistically modeling of breast microstructures [16,17]. These accomplishments enable the simulation of clinical trials on a per patient basis. These virtual clinical trials have become a feasible option for conducting preclinical testing of novel breast imaging systems [18]. The successes arising from simulation of the breast anatomy support the extension of octree-based recursive partitioning to the simulation of other septated tissues (e.g., cortical bone, lung parenchyma), as well

---

* Corresponding author.
  E-mail addresses: dexterofnis@gmail.com (M.D. Petković), Predrag.Bakic@uphs.upenn.edu (P.R. Bakic), Andrew.Maidment@uphs.upenn.edu (A.D.A. Maidment), dpokrajac@desu.edu (D. Pokrajac).

Notation

B	lattice (coordinate) box
K	number of partition functions
L	octree depth
ϕ_i	functions that partition box $B^{(0)}$ (shape functions)
$B^{(0)}$	initial cubic box
$B^{(l)}$	set of initial cubic box subdivisions with side length $\Delta^{(l)}$
B_{kl}	cell of a 2D lattice; projection of B onto xy plane
$\mathcal{B}^{(l)}$	set of all boxes $B^{(l)}$
$\dim_b(X)$	box-counting dimension of bounded set X
η	node of the octree
$\eta.\Phi$	set of shape functions associated with node η
$\eta.B$	cubic subdomain associated with node η
$\Delta^{(0)}$	side length of $B^{(0)}$
$\Delta^{(l)}$	$\Delta^{(0)}/2^l$ – side length of $B^{(l)}$
$\Phi^{(0)}$	set of shape functions
$\mathbb{Z}^3\Delta, \mathbb{Z}^2\Delta$	3D and 2D coordinate lattice with unit size Δ
p_L	total number of octree nodes up to level L
u_l	number of nodes at level l with subvolumes intersecting $\mathcal{S}$
v_l	number of nodes at level l of an octree
$C_k, k = 1, 2, \ldots, K$	subdomains of $B^{(0)}$, compartments
$N_X(\Delta), N_S(\Delta), N_\gamma(\Delta)$	number of boxes from $\mathbb{Z}^3\Delta$ intersecting with a bounded set $X, \mathcal{S}, \gamma$
$P(\mathcal{S})$	$N_S(\Delta)\Delta^2$
$\mathcal{C}^{(1)}(X)$	set of all continuously differentiable functions, defined on the set X
$\mu(X)$	lebesgue measure of the set X
$\overline{F}_{k,l}$	normalized maximal value of function f on $\mathcal{S}_{xy} \cap B_{kl}$
$F_{k,l}$	normalized minimal value of function f on $S_{xy} \cap B_{kl}$
R	union of rectangles R_i
R_i	rectangle on $\mathcal{S}_{xy}$
$\mathcal{S}$	union of all boundaries $\mathcal{S}_{ij}$; a surface
$\mathcal{S}_{ij}$	boundary of subdomains C_i and C_j
$\mathcal{S}_{xy}, \mathcal{S}_{xz}, \mathcal{S}_{yz}$	projections of surface S on coordinate planes; $\mathcal{S}_{xy}$ also denotes rectangle on plane xy
$\mathcal{S}_R$	surface induced by R
$M_{\mathcal{S}}$	$\max_{(x,y) \in \mathcal{S}_{xy}} \frac{\partial f}{\partial x}(x, y) + \max_{(x,y) \in \mathcal{S}_{xy}} \frac{\partial f}{\partial y}(x, y) + 2$
γ	finite union of $\mathcal{C}^{(1)}$ curves; A set with box-counting dimension <2
$\delta\mathcal{S}, \delta X$	boundaries of $\mathcal{S}, X$

as various porous materials [19]. To support the widespread application of this method and fully exploit the benefits, especially for multi-scale simulation tasks, a better understanding of its theoretical computational complexity is needed. This is the motivation for our work.

Note that the problem considered in [11] can be reduced ultimately to the problem of indexing voxels intersecting with a union of 3D $\mathcal{C}^{(1)}$ surfaces. The computational complexity of a recursive partitioning algorithm for approximation of 3D implicit *polynomial* surfaces was discussed in [20]. In that paper, using a concept of ϵ-entropy (the minimal number of closed balls of radius ϵ covering the surface), an upper bound for the algorithm complexity was proven. However, the problem of the lower bound of the problem, as well as the computational complexity of the similar, but not equivalent, problem discussed in [11] has remained open. Observe that experimental results [10,11] have indicated that the computational complexity of the recursive algorithm proposed to solve the problem is quadratic w.r.t. the reciprocal linear dimension of a voxel. This has led to the hypothesis that the asymptotic complexity of the algorithm is quadratic. Further, we wanted to examine the hypothesis that the algorithm is computationally optimal, which warrants determination of the lower bound of the complexity.

In this paper, we formally restate the considered problem and the recursive partitioning algorithm in Section 2. In Section 3, we demonstrate a quadratic computational complexity of the indexing algorithm. In addition, we demonstrate that the influence of the overhead introduced by recursive partitioning of the domain does not increase the asymptotic complexity of the algorithm. In Section 4, we formally prove the asymptotic number of uniform 3D cubic cells covering a finite union of $\mathcal{C}^{(1)}$ surfaces, which is the main result in Section 3.

2. Recursive partitioning indexing algorithm

Consider a cubic box $B^{(0)} \subset \mathbb{R}^3$ and functions $\phi_i \in \mathcal{C}^{(1)}(B^{(0)}), i = 1, 2, \ldots, K$ (hereafter referred to as *shape functions*). Define

$$\mathrm{md}(\mathbf{x}) = \min_{1 \leqslant j \leqslant K} \phi_j(\mathbf{x}).$$

The shape functions partition $B^{(0)}$ into K subdomains $C_1, C_2, \ldots, C_K$ defined by

$$C_k = \left\{ \mathbf{x} \in B^{(0)} \mid \phi_k(\mathbf{x}) = \mathrm{md}(\mathbf{x}) \right\}, \quad k = 1, 2, \ldots, K.$$

In other words, the set C_k is set of all points $\mathbf{x} \in B^{(0)}$ such that $\phi_k(\mathbf{x})$ is smallest among all values $\phi_j(\mathbf{x})$ for $j = 1, 2, \ldots, K$. These subdomains C_k are hereafter referred to as *compartments*. Note that compartments C_k provide a suitable generalization of Voronoi diagrams. Denote by $\mathcal{S}_{i,j}$ the common boundary of C_i and C_j ($1 \leqslant i < j \leqslant K$). Note that $\mathcal{S}_{i,j}$ may be an empty set if C_i and C_j do not have a common boundary. Furthermore, denote by $\mathcal{S}$ the union of all boundaries, i.e.,

$$\mathcal{S} = \bigcup_{1 \leqslant i < j \leqslant K} \mathcal{S}_{i,j} \tag{1}$$

It is obvious that all $\mathcal{S}_{i,j}$ are $\mathcal{C}^{(1)}$ surfaces implying that $\mathcal{S}$ is a piecewise $\mathcal{C}^{(1)}$ surface. We consider the following problem of indexing the boundaries of the compartments.

Assume that each side of the initial box $B^{(0)}$ with side length $\Delta^{(0)}$, is divided into a total 2^l equal parts. The entire box is therefore divided into a total of 2^{3l} subboxes of unit size $\Delta^{(l)} = \Delta^{(0)}/2^l$. Denote the set of all such subboxes by $\mathcal{B}^{(l)}$.

Problem 1 (**BoxApprox**). For a fixed $L \in \mathbb{N}$, find all subboxes $B \in \mathcal{B}^{(L)}$ of $B^{(0)}$ having non-empty intersection with $\mathcal{S}$. In other words, find all subboxes $B \in \mathcal{B}^{(L)}$ having non-empty intersection with at least two compartments C_k.

Note that $\mathcal{S}$ is a bounded surface. By an appropriate shift of the coordinate system, we obtain that all subboxes $B \in \mathcal{B}^{(L)}$ are coordinate boxes of a $\mathbb{Z}^3 \Delta^{(L)}$ coordinate lattice (i.e., a coordinate lattice with a unit size $\Delta^{(L)}$). Hence, we need to determine the set of all coordinate boxes B of $\mathbb{Z}^3 \Delta^{(L)}$ having non-empty intersection with $\mathcal{S}$.

A recursive partitioning algorithm to resolve this problem, which in addition indexes compartments, has been proposed [11]. The algorithm maintains an octree corresponding to $B^{(0)}$. Each node η at the level l of the octree is associated with a cubic subdomain $\eta.B \in \mathcal{B}^{(l)}$ of interest. Also, for each node the set $\eta.\Phi$ of shape functions is kept such that

$$\eta.\Phi = \{\phi_i \mid \phi_i(\mathbf{x}) = \mathrm{md}(\mathbf{x}), \text{ for some } \mathbf{x} \in \eta.B\}.$$

in a breadth-first fashion [21]. Nodes at each level of the tree are successively examined; if $\eta.B$ does not intersect $\mathcal{S}$ (i.e., $\eta.\Phi$ is a one element set) the node is not further split. Otherwise, the node is split into eight nodes $\eta_1, \eta_2, \ldots, \eta_8$ of the subsequent tree level $l + 1$ and the corresponding values $\eta_k.B$ and $\eta_k.\Phi$ are determined. The recursive partitioning procedure continues until the tree depth L is reached. The algorithm from [11] can be conceptualized using the pseudocode notation, shown in Algorithm 2.1. Note that the function *SplitVolume* extracts a kth subvolume $\eta_k.B$ from a given volume $\eta.B$. The function *RefineShapeFunctions* determine which shape functions $\eta_k.\Phi$ are associated to a subvolume $\eta_k.B$, based on the shape functions $\eta.\Phi$ associated to a node η.

Algorithm 2.1. RecursivePartitionIndexing($B^{(0)}; \Phi^{(0)}; L; K$)

Require: Root volume $B^{(0)}$, shape functions $\Phi^{(0)} = \{\phi_1(\mathbf{x}), \phi_2(\mathbf{x}), \ldots, \phi_K(\mathbf{x})\}$; and the depth L of the octree
1: $Root.B := B^{(0)}$
2: $Root.\Phi := \Phi^{(0)}$
3: **for** level $l := 0$ to L-1 **do**
4:    **for** each octree node η at level l
5:       **if** $|\eta.\Phi| > 1$ **then**
6:          Split the node η into identical-sized, level $l + 1$ subnodes $\eta_k, k = 1, \ldots, 8$
7:          **for** $k := 1$ to 8 **do**
8:             $\eta_k.B := SplitVolume(\eta.B, k)$
9:             $\eta_k.\Phi := RefineShapeFunctions(\eta_k.B, \eta.\Phi)$
10:         **end for**
11:       **end if**
12:    **end for**
13: **end for**
14: **for** each leaf node η in the tree **do**
15:    **return** $\eta.B, \eta.\Phi$
16: **end for**

---

3. Computational complexity

In this section, we provide the computational complexity of Problem 1 and Algorithm 2.1. Denote by $N_X(\Delta)$ the number of boxes from lattice $\mathbb{Z}^3 \Delta$ (i.e., a coordinate lattice with unit size Δ) having non-empty intersection with bounded set $X \subset \mathbb{R}^3$.

3.1. Asymptotic complexity of problem 1

The computational complexity of Problem 1 is determined by the number $N_S(\Delta^{(L)})$ of cubes $B^{(L)} \in \mathcal{B}^{(L)}$. Recall that $\Delta^{(L)} = \Delta^{(0)}/2^L$ where $\Delta^{(0)}$ is the size of initial bounding box $B^{(0)}$.

The following theorem, utilized in the rest of this Section, is proven in Section 4.

Theorem 3.1. *Assume that $S \subset \mathbb{R}^3$ is a bounded, piecewise $C^{(1)}$ surface. Then $N_S(\Delta)\Delta^2 \to const$ when $\Delta \to 0$.*

The Theorem 3.2 below directly gives the asymptotic computational complexity of the Problem 1.

Theorem 3.2. *The asymptotic computational complexity of Problem 1 is $\Theta(2^{2L})$. In other words, the number $N_S(\Delta^{(L)})$ of all subboxes $B^{(L)} \in \mathcal{B}^{(L)}$ having non-empty intersection with S is asymptotically $\Theta(2^{2L})$.*

Proof. It is easy to see that S defined by (1) is piecewise $C^{(1)}$ surface, since all $\phi_i(\mathbf{x}), i = 1, 2, \ldots, K$ are $C^{(1)}$ functions. The statement of the theorem now follows from Theorem 3.1. □

3.2. Asymptotic complexity of recursive partitioning algorithm

Denote by v_l, the number of nodes at level l of the tree. Let u_l denote the number of nodes intersecting the surface S at level l of the tree. Denote by p_L, the total number of nodes of the octree on levels $1, 2, \ldots, L$. As demonstrated in [11], function *SplitVolume* has $\mathcal{O}(1)$ complexity. Also, function *RefineShapeFunctions* has $\mathcal{O}(K)$ complexity. Hence, the computational complexity of the Algorithm 2.1 is $\mathcal{O}(p_L)$.

The number of nodes u_l in the level l of the tree containing the compartment boundaries is equal to the number $N_S(\Delta^{(l)})$ of cubes with linear dimension $\Delta^{(l)} = \Delta^{(0)}/2^l$ containing the boundaries. Fig. 1 visualizes the octree we are considering. Black nodes depict the tree nodes containing compartment boundaries (intersecting the surface S). Gray nodes contain only compartments. Note that u_l is the number of black, while v_l is the number of both black and gray nodes at level l.

Our main tools in the proof of an asymptotic formula for p_L are Corollary 3.1 (following directly from Theorem 3.1) and Lemma 3.1.

Corollary 3.1. *There exists a constant C and a sequence $\omega_l \to 0$ $(l \to +\infty)$ such that $u_l/4^l = C + \omega_l$.*

Lemma 3.1. *Assume that δ_l is any sequence such that $\delta_l \to 0$ when $l \to +\infty$. Denote by*

$$\gamma_L = \frac{\delta_{L-1}}{4} + \frac{\delta_{L-2}}{4^2} + \cdots + \frac{\delta_0}{4^L}.$$

Then $\gamma_L \to 0$ when $L \to +\infty$.

Proof. Fix $\epsilon > 0$ and denote by L' the number such that $|\delta_l| < \epsilon$ for every $l > L'$. For each $L > L'$ holds
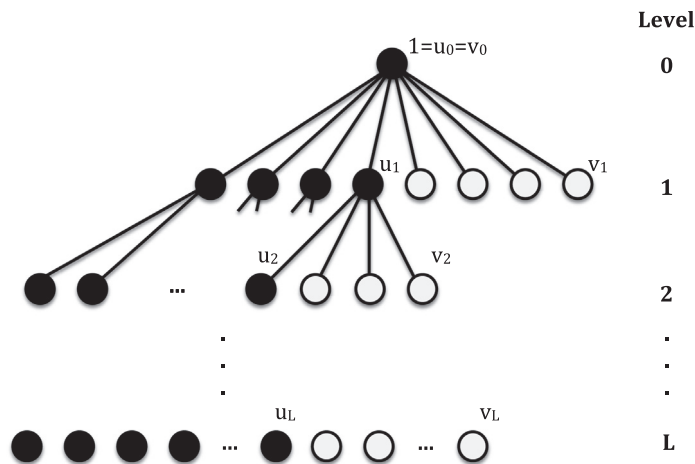
Fig. 1. Visualization of the octree. Black nodes depict the tree nodes containing compartment boundaries (intersecting the surface S). Gray nodes contain only compartments. The number of black nodes at level l and the total number of nodes at the level l are denoted with u_l and v_l, respectively.

$$|\gamma_L| \leqslant \left| \frac{\delta_{L-1}}{4} + \frac{\delta_{L-2}}{4^2} + \cdots + \frac{\delta_{L'}}{4^{L-L'}} \right| + \left| \frac{\delta_{L'}}{4^{L-L'}} + \frac{\delta_{L'-1}}{4^{L-LL'+1}} + \cdots + \frac{\delta_0}{4^L} \right|$$

$$\leqslant \epsilon \left(\frac{1}{4} + \frac{1}{4^2} + \cdots + \frac{1}{4^{L-L'+1}} \right) + \frac{1}{4^{L-L'}} \left| \delta_{L'} + \frac{\delta_{L'-1}}{4} + \cdots + \frac{\delta_0}{4^{L'}} \right|$$

$$= \epsilon \frac{1 - 4^{-(L-L'+1)}}{3} + 4^{-(L-L')} h(L')$$

where

$$h(L') = \left| \sum_{l=0}^{L'} 4^{-l} \delta_{L'-l} \right|$$

does not depend on L. It is possible to choose sufficiently large L, so that $4^{-(L-L')} h(L') < 2\epsilon/3$ and subsequently $|\gamma_L| < \epsilon$. This completes the proof of lemma. $\quad\square$

However, p_L is directly proportional to the computation complexity of Algorithm 2.1 and equal to

$$p_L = v_0 + v_1 + \cdots + v_L = 1 + 8(u_0 + u_1 + \cdots + u_{L-1}).$$

According to Corollary 3.1:

$$p_L = 1 + 8C(4^0 + 4^1 + \cdots + 4^{L-1}) + 8(\omega_0 + \omega_1 4^1 + \cdots + \omega_{L-1} 4^{L-1})$$

and hence

$$\frac{p_L}{4^L} = 4^{-L} + 8C \frac{1 - 4^{-L}}{3} + \frac{\omega_{L-1}}{4} + \frac{\omega_{L-2}}{4^2} + \cdots + \frac{\omega_0}{4^{L-1}}.$$

According to Lemma 3.1 we get

$$\lim_{L \to +\infty} \left| \frac{p_L}{4^L} - 8C \frac{1 - 4^{-L}}{3} \right| = 0$$

implying that

$$\lim_{L \to +\infty} \frac{p_L}{4^L} = \frac{8C}{3} \quad \Rightarrow \quad p_L = \Theta(4^L). \tag{2}$$

According to this subsection (especially (2)), the following theorem holds:

Theorem 3.3. *The computational complexity of Algorithm 2.1 is $\Theta(4^L)$.*

4. Asymptotic number of Z^3 cells covering a piecewise $\mathcal{C}^{(1)}$ surface

Formal proof of Theorem 3.1 is provided stepwise in this section. From this point forward, let $\mathcal{S}$ represent an arbitrary (piecewise) $\mathcal{C}^{(1)}$ surface and $\Delta > 0$ is a real number.

4.1. Statement of the problem

Let $\mathcal{S} \subset \mathbb{R}^3$ be the surface and denote by $\mathcal{S}_{xy}, \mathcal{S}_{yz}$ and $\mathcal{S}_{xz}$ its projections to the xy, yz and xz coordinate planes. Assume that $\mathcal{S}$ can be represented as the graph of the function defined on $\mathcal{S}_{xy}$, i.e.,

$$\mathcal{S} = \{ (x, y, f(x, y)) \mid (x, y) \in \mathcal{S}_{xy} \} \tag{3}$$

where $f : \mathcal{S}_{xy} \to \mathbb{R}$ is $\mathcal{C}^{(1)}$ function and $\mathcal{S}_{xy}$ is compact. Also assume that the boundary of $\mathcal{S}$ is a finite length curve.

Recall that, by $N_X(\Delta)$ we denote the number of boxes from lattice $\mathbb{Z}^3 \Delta$ (i.e., a coordinate lattice with unit size Δ) which have non-empty intersection with bounded set $X \subset \mathbb{R}^3$. Also for the bounded set $X \subset \mathbb{R}^2$, denote by $\mu(X)$ the (Lebesgue) measure of the set X. Further, denote by $N_X(\Delta)$ the number of squares from corresponding $\mathbb{Z}^2 \Delta$ lattice in a plane that has non-empty intersection with $X \subset \mathbb{R}^2$.

The following definition of the box-counting dimension $\dim_b(X)$ of the bounded set $X \subset \mathbb{R}^3$ is well-known (see for example [22]):

$$\dim_b(X) = \lim_{\Delta \to 0} -\frac{\log N_X(\Delta)}{\log \Delta}. \tag{4}$$

It is also well-known (see for example [22]) that the box-counting dimensions of the surface $\mathcal{S}$ is equal to $\dim_b(\mathcal{S}) = 2$. This means that for each $\epsilon > 0$, there is $\Delta_\epsilon > 0$ such that $N_\mathcal{S}(\Delta)\Delta^2 \in (\Delta^\epsilon, \Delta^{-\epsilon})$ holds for each $\Delta < \Delta_\epsilon$. However, it does not automatically imply that $N_\mathcal{S}(\Delta)\Delta^2$ converges to some constant C (when $\Delta \to 0$). Even more, it does not even imply that $N_\mathcal{S}(\Delta)\Delta^2$ is bounded either from the top or from the bottom. In the rest of this Section we show that $N_\mathcal{S}(\Delta)\Delta^2$ converges (when $\Delta \to 0$) and we find its limit value.

4.2. Preliminary results

Assume that $\mathcal{S}$ does not pass through any node of lattice $\mathbb{Z}^3\Delta$. Without loss of generality, assume that $\mathcal{S}_{xy}$ lies in the first quadrant of the xy coordinate plane. Denote by $B_{kl} = [k\Delta, (k + 1)\Delta] \times [l\Delta, (l + 1)\Delta]$ and

$$\bar{F}_{k,l} = \left\lceil \max_{(x,y) \in \mathcal{S}_{xy} \cap B_{kl}} f(x,y)/\Delta \right\rceil, \quad F_{k,l} = \left\lfloor \min_{(x,y) \in \mathcal{S}_{xy} \cap B_{kl}} f(x,y)/\Delta \right\rfloor$$

for every $k, l \in \mathbb{N}_0$ such that B_{kl} have non-empty intersection with $\mathcal{S}_{xy}$. The following propositions directly follow from continuity of f.

Proposition 4.1. *The number of coordinate boxes with the base B_{kl} having non-empty intersection with $\mathcal{S}$ is equal to $\bar{F}_{k,l} - F_{k,l}$. Hence,*

$$N_{\mathcal{S}}(\Delta) = \sum_{B_{kl} \cap \mathcal{S}_{xy} \neq \emptyset} (\bar{F}_{k,l} - F_{k,l}). \tag{5}$$

Proposition 4.2. *If $X \subset \mathbb{R}^2$ is compact then*

$$\lim_{\Delta \to 0} N_X(\Delta)\Delta^2 = \mu(X).$$

Proposition 4.3. *If $\gamma \subset \mathbb{R}^3$ is any set with box-counting dimension less than 2, then $N_\gamma(\Delta)\Delta^2 \to 0$ when $\Delta \to 0$.*

Proof. Assume that $d = \dim_b(\gamma)$. According to (4), we get $N_\gamma(\Delta) \leqslant \Delta^{-d-\epsilon}$ for some $\epsilon < 2 - d$ and $\Delta < \Delta_\epsilon$. Hence

$$N_\gamma(\Delta)\Delta^2 \leqslant \Delta^{2-d-\epsilon} \to 0, \qquad \Delta \to 0.$$

This completes the proof. □

Corollary 4.1. *If $\gamma \subset \mathbb{R}^3$ is a finite union of $\mathcal{C}^{(1)}$ curves, then $N_\gamma(\Delta)\Delta^2 \to 0$ when $\Delta \to 0$.*

Proof. It is well-known (see for example [22]) that $\dim_b(\gamma) = 1$. Now the statement of the corollary follows directly from the previous proposition. □

Denote by $\delta\mathcal{S}$ the boundary of a surface $\mathcal{S}$ and by δX the boundary of a compact set $X \subset \mathbb{R}^2$. In what follows, we always assume that the boundary of every surface is a piecewise $\mathcal{C}^{(1)}$ curve.

Proposition 4.4. *If $\mathcal{S} = \mathcal{S}^1 \cup \mathcal{S}^2 \cup \cdots \cup \mathcal{S}^v$, where $\mathcal{S}^i$ and $\mathcal{S}^j$ may have only boundary points in common (in other words $\mathcal{S}^i \cap \mathcal{S}^j \subset \delta\mathcal{S}^i \cup \delta\mathcal{S}^j$) then*

$$0 \leqslant \sum_{i=1}^{v} N_{\mathcal{S}^i}(\Delta) - N_{\mathcal{S}}(\Delta) \leqslant (v - 1)N_\gamma(\Delta)$$

where $\gamma = \cup_{i=1}^{v} \delta\mathcal{S}^i$. If $N_{\mathcal{S}^i}(\Delta)\Delta^2 \to L_i$ when $\Delta \to 0$ for every $i = 1, 2, \ldots, v$, then $N_{\mathcal{S}}(\Delta)\Delta^2 \to \sum_{i=1}^{v} L_i$ when $\Delta \to 0$.

Proof. Denote by $s = \sum_{i=1}^{v} N_{\mathcal{S}^i}(\Delta)$. Assume that a coordinate box B intersects $\mathcal{S}$ but not γ. Then, there is exactly one i such that B intersects $\mathcal{S}_i$. Those boxes are counted once in s. On the other hand, if B intersects γ, it might intersect more than one surface $\mathcal{S}_i$. Those boxes are counted at most v times in s. Therefore, the difference

$$\sum_{i=1}^{v} N_{\mathcal{S}^i}(\Delta) - N_{\mathcal{S}}(\Delta)$$

gives the number of all additional counts of boxes, which are at most $(v - 1)N_\gamma(\Delta)$.

The second part follows from the fact that $\delta\mathcal{S}_i$ is a piecewise $\mathcal{C}^{(1)}$ curve, implying that γ is finite union of such curves and $N_\gamma(\Delta)\Delta^2 \to 0$ according to Corollary 4.1. □

Denote by $N_{\mathcal{S}}^{int}(\Delta)$ the number of boxes B which intersect $\mathcal{S}$ but not $\delta\mathcal{S}$. In the same sense, denote by $N_{\mathcal{S}}^{\delta}(\Delta) = N_{\delta\mathcal{S}}(\Delta)$ the number of boxes having non-empty intersection with $\delta\mathcal{S}$. It is obvious that $N_{\mathcal{S}}(\Delta) = N_{\mathcal{S}}^{int}(\Delta) + N_{\mathcal{S}}^{\delta}(\Delta)$. Moreover, if B intersects $\mathcal{S}$ but not $\delta\mathcal{S}$, then its projection B_{kl} on the xy plane intersects $\mathcal{S}_{xy}$ but not $\delta\mathcal{S}_{xy}$. Therefore $B_{kl} \subset \int \mathcal{S}_{xy}$ and from Proposition 4.1 it holds that

$$N_{\mathcal{S}}^{int}(\Delta) = \sum_{B_{kl} \subset \text{int}\mathcal{S}_{xy}} (\bar{F}_{k,l} - F_{k,l}). \tag{6}$$

Lemma 4.1. *Assume that $\mathcal{S}_{xy}$ is compact with non-empty interior and denote by*

$$M_{\mathcal{S}} = \max_{(x,y)\in\mathcal{S}_{xy}} \frac{\partial f}{\partial x}(x,y) + \max_{(x,y)\in\mathcal{S}_{xy}} \frac{\partial f}{\partial y}(x,y) + 2. \tag{7}$$

Then

$$N_{\mathcal{S}}^{int}(\Delta) \leqslant M_{\mathcal{S}} \cdot N_{\mathcal{S}_{xy}}(\Delta).$$

Proof. Assume that $B_{kl} \subset \mathcal{S}_{xy}$. Since f is continuous and B_{kl} is compact, there exist points $(x_0, y_0), (x_1, y_1) \in B_{kl} \cap \mathcal{S}_{xy}$ where $f(x, y)$ attains its minimum and maximum values, respectively, i.e., $F_{k,l} = \lfloor f(x_0, y_0)/\Delta \rfloor$ and $\bar{F}_{k,l} = \lceil f(x_1, y_1)/\Delta \rceil$. According to the Lagrange theorem, there exists (x', y') on the line connecting (x_0, y_0) and (x_1, y_1), such that

$$\begin{aligned}
\bar{F}_{k,l} - F_{k,l} &\leqslant \tfrac{f(x_1,y_1)-f(x_0,y_0)}{\Delta} + 2 \\
&\leqslant [\tfrac{\partial f}{\partial x}(x',y')\tfrac{x_1-x_0}{\Delta} + \tfrac{\partial f}{\partial y}(x',y')\tfrac{y_1-y_0}{\Delta}] + 2 \\
&\leqslant \tfrac{\partial f}{\partial x}(x',y') + \tfrac{\partial f}{\partial y}(x',y') + 2 \leqslant M_{\mathcal{S}}.
\end{aligned}$$

The number of boxes $B_{kl} \subset \mathcal{S}_{xy}$ is less than or equal to $N_{\mathcal{S}_{xy}}(\Delta)$ and hence it is valid:

$$N_{\mathcal{S}}^{int}(\Delta) \leqslant M_{\mathcal{S}} N_{\mathcal{S}_{xy}}(\Delta).$$

This completes the proof. □

4.3. Main result

Assume now that $\mathcal{S}$ can be represented in the same way as in (3), but taking $\mathcal{S}_{xz}$ and $\mathcal{S}_{yz}$ respectively as the domain of the function f. In other words, assume that

$$\mathcal{S} = \{(x, f_{xz}(x,z), z)|(x,z) \in \mathcal{S}_{xz}\} = \{(f_{yz}(y,z), y, z)|(y,z) \in \mathcal{S}_{yz}\}.$$

It can be easily seen that $f(\bar{x}, \cdot)$ and $f(\cdot, \bar{y})$ are bijections. Indeed, if $z = f(\bar{x}, y_0) = f(\bar{x}, y_1)$, for some $(\bar{x}, y_0), (\bar{x}, y_1) \in \mathcal{S}_{xy}$, then $(\bar{x}, y_0, \bar{z}) \in \mathcal{S}$ and $(\bar{x}, y_1, \bar{z}) \in \mathcal{S}$ which implies that $y_0 = y_1 = f_{xz}(\bar{x}, \bar{z})$. Assume that $\partial f/\partial x(x', y') = 0$ for some $(x', y') \in \mathrm{int}\mathcal{S}_{xy}$. Consider the function $f(\cdot, y')$ in the neighborhood of (x', y'). It has a local minimum in x' which means that it is not a bijection. Therefore, $\partial f/\partial x \neq 0$ and hence it does not change its sign on $\mathcal{S}_{xy}$. The same holds for $\partial f/\partial y$. Without loosing generality, we may assume that both partial derivatives are positive on $\mathcal{S}_{xy}$.

In what follows, we show the proof of the following theorem.

Theorem 4.1. *The following is valid*

$$\lim_{\Delta\to 0} N_{\mathcal{S}}(\Delta)\Delta^2 = P(\mathcal{S}) = \mu(\mathcal{S}_{xy}) + \mu(\mathcal{S}_{yz}) + \mu(\mathcal{S}_{xz}) \tag{8}$$

where the limit is taken over those values Δ such that $\mathcal{S}$ do not contain any node of the lattice $\mathbb{Z}^3\Delta$.

First we prove the theorem for rectangular $\mathcal{S}_{xy}$.

Lemma 4.2. *Assume that $\mathcal{S}_{xy}$ is rectangle $[x_{min}, x_{max}] \times [y_{min}, y_{max}]$. Then (8) is valid.*

Proof. Without loss of generality, assume that $x_{min} \in [0, \Delta)$. If the latter is not true, then translate the surface by the vector $(\lfloor x_{min}/\Delta \rfloor \Delta, 0, 0)$. In the same way, assume that $y_{min} \in [0, \Delta)$. Let $m = \lfloor x_{max}/\Delta \rfloor$ and $n = \lfloor y_{max}/\Delta \rfloor$. Furthermore, define

$$x_k = \begin{cases} x_{min}, & k = 0 \\ k\Delta, & k = 1, 2 \dots, m \\ x_{max}, & k = m+1. \end{cases}$$

and analogously y_l for $l = 0, 1, \dots, n+1$. According to Proposition 4.1, we have

$$N_{\mathcal{S}}(\Delta) = \sum_{k=0}^{m}\sum_{l=0}^{n}(\bar{F}_{k,l} - F_{k,l}). \tag{9}$$

Note that $B_{kl} \subset \mathcal{S}_{xy}$ for $k = 1, 2, \dots, m-1$ and $l = 1, 2, \dots, n-1$. Since $f(x, y)$ is monotonically increasing function of both x and y, we conclude that

$$\begin{aligned}
F_{k,l} &= \lfloor f(x_k, y_l)/\Delta \rfloor \\
\bar{F}_{k,l} &= \lceil f(x_{k+1}, y_{l+1})/\Delta \rceil = F_{k+1,l+1} + 1 \\
k &= 1, \dots, m-1, \\
l &= 1, \dots, n-1.
\end{aligned}$$

Replacing this in (9) yields

$$N_S(\Delta) = mn + \sum_{k=0}^{m}(\bar{F}_{k,n} - F_{k,0}) + \sum_{l=0}^{n}(\bar{F}_{m,l} - F_{0,l}) - (\bar{F}_{m,n} - F_{0,0}).$$ (10)

Now consider S_{xz}. Due to the continuity and monotonicity of f, it can be written in the form

$$S_{xz} = \{(x,z)|x \in [x_{min}, x_{max}], \ z \in [f(x, y_{min}), f(x, y_{max})]\}.$$

Thus, it is evident that the number of boxes of $\mathbb{Z}^2\Delta$ having non-empty intersection with $S_{xz} \cap ([x_k, x_{k+1}] \times \mathbb{R})$ is equal to

$$\bar{F}_{k,n} - F_{k,0} = \lceil f(x_{k+1}, y_{max})/\Delta \rceil - \lfloor f(x_k, y_{max})/\Delta \rfloor.$$

Therefore, the first sum in (10) is equal to $N_{S_{xz}}(\Delta)$. The same way, the second sum is $N_{S_{yz}}(\Delta)$. According to these, we now have

$$N_S(\Delta)\Delta^2 = mn\Delta^2 + N_{S_{xz}}(\Delta)\Delta^2 + N_{S_{yz}}(\Delta)\Delta^2 - (\bar{F}_{m,n} - F_{0,0})\Delta^2.$$ (11)

Since $|x_{min}| < \Delta$ (by assumption from the beginning of the proof), we get

$$|x_{max} - x_{min} - m\Delta| \leqslant |x_{max} - m\Delta| + \Delta \leqslant 2\Delta$$

and similarly $|y_{max} - y_{min} - n\Delta| \leqslant 2\Delta$. Hence, the limit of the first term in (11) is $\mu(S_{xy}) = (x_{max} - x_{min})(y_{max} - y_{min})$, when $\Delta \to 0$. Proposition 4.2 implies that the second and third term in (11) tend to $\mu(S_{xz})$ and $\mu(S_{yz})$ respectively. Finally, the forth term can be bounded as

$$(\bar{F}_{m,n} - F_{0,0})\Delta^2 \leqslant \left[\frac{f(x_{max}, y_{max})}{\Delta} - \frac{f(x_{min}, y_{min})}{\Delta} + 2\right]\Delta^2$$

and hence tends to 0. This completes the proof of the Lemma 4.2. □

Proof (*Proof of the main theorem*). Since S_{xy} is compact, it can be approximated by the union of rectangles $R_i \subseteq S_{xy}, i = 1, 2, \ldots, \nu$ having only boundary points in common, such that

$$\mu(\bar{S}_{xy}) < \frac{\epsilon}{12M_S}, \quad \bar{S}_{xy} = S_{xy} \setminus R, \quad R = \cup_{i=1}^{\nu} R_i$$

and $P(S) - P(S_R) < \epsilon/3$, where $\epsilon > 0$ is arbitrary and

$$S_R = \{(x, y, f(x, y))|(x, y) \in R\}$$

is a surface induced by the union of the rectangles R. The same way, define S_{R_i} and S_γ where $\gamma = \cup_{i=1}^{\nu}\delta S_{R_i} \cup \delta S$. Furthermore, let $\bar{S} = S \setminus S_R$. Note that $\delta\bar{S} = \delta S_R \cup \delta S \subseteq \gamma$. According to Lemma 4.2 and Proposition 4.4, there exist $\Delta_1 > 0$ such that

$$\left|N_{S_R}(\Delta)\Delta^2 - P(S_R)\right| \leqslant \frac{\epsilon}{3}$$

for all $\Delta \leqslant \Delta_1$. Furthermore, according to Proposition 4.4 we have

$$|N_S(\Delta)\Delta^2 - N_{S_R}(\Delta)\Delta^2| \leqslant N_{\bar{S}}(\Delta)\Delta^2 + \nu N_\gamma(\Delta)\Delta^2 \leqslant N_{\bar{S}}^{int}(\Delta)\Delta^2 + (\nu + 1)N_\gamma(\Delta)\Delta^2.$$

The last inequality holds since $\delta\bar{S} \subseteq \gamma$ and hence $N_{\bar{S}}^\delta(\Delta) = N_{\delta\bar{S}}(\Delta) \leqslant N_\gamma(\Delta)$. The first term can be bounded (Lemma 4.1) by:

$$N_{\bar{S}}^{int}(\Delta)\Delta^2 \leqslant M_{\bar{S}} \cdot N_{\bar{S}_{xy}}(\Delta)\Delta^2 \leqslant M_S \cdot N_{\bar{S}_{xy}}(\Delta)\Delta^2.$$

Here we used that $M_{\bar{S}} \leqslant M_S$ since the maximum is taken over the larger set (see (7)). According to Proposition 4.2, we can choose $\Delta_2 > 0$ such that $|N_{\bar{S}_{xy}}(\Delta)\Delta^2 - \mu(\bar{S}_{xy})| \leqslant \epsilon/(12M_S)$ for all $\Delta \leqslant \Delta_2$. Then

$$|N_{\bar{S}_{xy}}(\Delta)\Delta^2| \leqslant \frac{\epsilon}{12M_S} + \mu(\bar{S}_{xy}) \leqslant \frac{\epsilon}{6M_S}$$

and

$$N_{\bar{S}}^{int}(\Delta)\Delta^2 \leqslant M_{\bar{S}}\frac{\epsilon}{6M_S} \leqslant \frac{\epsilon}{6}.$$

Corollary 4.1 yields that there exists $\Delta_3 > 0$ such that $N_\gamma(\Delta)\Delta^2 < \epsilon/(6(\nu + 1))$ for all $\Delta \leqslant \Delta_3$. Putting everything together, we get

$$|N_S(\Delta)\Delta^2 - N_{S_R}(\Delta)\Delta^2| \leqslant \frac{\epsilon}{3}.$$

Finally

$$|N_S(\Delta)\Delta^2 - P(S)| \leqslant |N_S(\Delta)\Delta^2 - N_{S_R}(\Delta)\Delta^2| + |N_{S_R}(\Delta)\Delta^2 - P(S_R)| + |P(S_R) - P(S)| \leqslant \epsilon$$

for every $\Delta < \min\{\Delta_1, \Delta_2, \Delta_3\}$. This completes the proof of the theorem. □

4.4. Extensions on arbitrary (piecewise) $\mathcal{C}^{(1)}$ surfaces

Theorem 4.1 implies that $N_{\mathcal{S}}(\Delta)\Delta^2$ tends to $P(\mathcal{S})$ when $\Delta \to 0$, if $\mathcal{S}$ is a graph surface of the $\mathcal{C}^{(1)}$ function f defined on the compact $\mathcal{S}_{xy}$. Moreover, assumption is that $\mathcal{S}$ can be also represented as the graph surface on $\mathcal{S}_{yz}$ and $\mathcal{S}_{xz}$.

Note that any $\mathcal{C}^{(1)}$ surface can be divided on finitely many graph surfaces. Also, each graph $\mathcal{C}^{(1)}$ surface can be divided on finitely many surfaces that can be represented as graph surfaces on all three projections.

In that sense, every $\mathcal{C}^{(1)}$ surface $\mathcal{S}$ can be divided on finitely many $\mathcal{C}^{(1)}$ surfaces $\mathcal{S}^i, i = 1, 2, \ldots, m$ satisfying the conditions assumed in the previous section. According to Proposition 4.4 and Theorem 4.1, we conclude that

$$\lim_{\Delta \to 0} N_{\mathcal{S}}(\Delta)\Delta^2 = \sum_{i=1}^{m} P(\mathcal{S}^i).$$

Therefore, the following Corollary holds, which is equivalent to Theorem 3.1.

Corollary 4.2. *Let $\mathcal{S}$ be a union of a finite number of $\mathcal{C}^{(1)}$ surfaces. Then,*

$$\lim_{\Delta \to 0} N_{\mathcal{S}}(\Delta)\Delta^2 = C$$

where C is a constant.

5. Discussion and conclusion

In this paper, we considered the problem of indexing cells on a uniform grid intersecting with a union of a finite number of $\mathcal{C}^{(1)}$ surfaces. As a main result, we prove that the problem has quadratic asymptotic complexity in the reciprocal linear size of a grid cell. Also, we demonstrated that a practical recursive partitioning algorithm [11] that indexes the grid cells intersecting with the surfaces achieves the problem complexity bound and therefore is asymptotically optimal. We believe this opens the venue for further application of the algorithm in multi-scale simulation.

Note that in [11], a statistical analysis of the algorithm complexity was performed. There, execution time of multiple runs of the algorithm implementation was regressed as a function of a power of a reciprocal linear size of a grid cell. Using the t-test, the hypothesis that the power coefficient of the regression model is 2 could not be rejected with significance $\alpha = 0.05$ (see e.g., [23] for details on regression models estimation and inference). These results are consistent with the theoretical considerations presented in this paper.

Work continues on generalization of the main result when the indexed surface has a box-counting dimension different from 2.

References

[1] D. Meagher, Geometric modeling using octree encoding, Comput. Graphics Image Process. 19 (1982) 129–147.
[2] H. Samet, The quadtree and related hierarchical data structures, ACM Comput. Surv. (CSUR) 16 (2) (1984) 187–260.
[3] H. Samet, A. Kochut, Octree approximation and compression methods, in: Proceedings First International Symposium on 3D Data Processing Visualization and Transmission, 2002, pp. 460–469.
[4] J. Wilhelms, A. Van Gelder, Octrees for faster isosurface generation, IEEE Trans. Med. Imaging 19 (2000) 739–758.
[5] W. Song, S. Hua, Z. Ou, H. An, Octree based representation and volume rendering of three-dimensional medical data sets, in: Proceedings International Conference on BioMedical Engineering and Informatics, BMEI, 2008, pp. 316–320.
[6] A. Badal, I. Kyprianou, A. Badano, K.J. Myers, J. Sempau, Monte Carlo package for simulating radiographic images of realistic anthropomorphic phantoms described by triangle meshes, Medical Imaging 2007, Proc. SPIE 6510 (2007) 65100Z.
[7] W.-T. Wong, F.Y. Shih, T.-F. Su, Thinning algorithms based on quadtree and octree representations, Inf. Sci. 176 (2006) 1379–1394.
[8] H. Lee, M. Desbrun, P. Schöder, Progressive encoding of complex isosurfaces, in: Proceedings ACM SIGGRAPH 2003 Conference, 2003, pp. 471–476.
[9] K. Zhou, M. Gong, X.H. Baining, Data-parallel octrees for surface reconstruction, IEEE Trans. Visual. Comput. Graphics 17 (5) (2011) 669–681.
[10] J.H. Chui, D.D. Pokrajac, A.D.A. Maidment, P.R. Bakic, Towards breast anatomy simulation using GPUs, in: Proceedings 11th International Workshop on Breast Imaging (IWDM 2012), Lecture Notes in Computer Science 7361, 2012, pp. 506–513.
[11] D.D. Pokrajac, A.D.A. Maidment, P.R. Bakic, Optimized generation of high resolution breast anthropomorphic software phantoms, Med. Phys. 39 (2012) 2290–2302.
[12] J. Chui, D. Pokrajac, A.D. Maidment, P. Bakic, Roadmap for Efficient Parallelization of Breast Anatomy Simulation, Medical Imaging 2012, Proc. SPIE 8313, 2012, 83134T.
[13] F. Chen, D. Pokrajac, F. Liu, A.D. Maidment, P. Bakic, Partial volume effect simulation in software breast phantoms, in: cal Imaging 2012, Proc. SPIE 8313, 2012, 83134U.

[14] F. Chen, D. Pokrajac, X. Shi, F. Liu, A.D. Maidment, P. Bakic, Simulation of three materials partial volume averaging in a software breast phantom, in: Proc. 11th International Workshop on Breast Imaging (IWDM), Lecture Notes in Computer Science 7361, 2012, pp. 151-158.
[15] J. Chui, R. Zeng, D. Pokrajac, S. Park, K.J. Myers, A.D.A. Maidment, P.R. Bakic, Two methods for simulation of dense tissue distribution in software breast phantoms, in: Medical Imaging 2013, Proc. SPIE 8668, 2013, 86680M.
[16] C.K. Abbey, P.R. Bakic, D.D. Pokrajac, A.D.A. Maidment, M.P. Eckstein, J.M. Boone, Non-gaussian statistical properties of virtual breast phantoms, in: Medical Imaging 2014: Image Perception, Observer Performance, and Technology Assessment, Proc. SPIE 9037, 2014, 90370G.
[17] P.R. Bakic, D.D. Pokrajac, R. DeCaro, A.D.A. Maidment, Realistic Simulation of Breast Tissue Microstructure in Software Anthropomorphic Phantoms. in: Proceedings 12th International Workshop on Breast Imaging (IWDM), Lecture Notes in Computer Science 8539, pp. 348–355.
[18] A.D.A. Maidment, Virtual clinical trials for the assessment of novel breast screening modalities, in: Proceedings 12th International Workshop on Breast Imaging (IWDM), Lecture Notes in Computer Science 8539, 2014, pp. 1–8.
[19] P. Liu, G.-F. Chen, Porous Materials: Processing and Applications, Butterworth-Heinemann, Oxford, UK, 2014.
[20] L. Alberti, G. Comte, B. Mourrain, Meshing implicit algebraic surfaces: the smooth case, in: Proceedings Conference on Mathematical Methods for Curves and Surfaces, 2005, pp. 11–26.
[21] E.F. Moore, The shortest path through a maze, in: Proceedings of the International Symposium on the Theory of Switching, 1959, pp. 285–292.
[22] K. Falconer, Fractal geometry – Mathematical Foundations and Applications, second edition., Wiley, 2003.
[23] J.L. Devore, Probability and Statistics for Engineering and the Sciences, Brooks/Cole, Belmont, CA, 2008.

Description and Characterization of a Novel Method for
Partial Volume Simulation in Software Breast Phantoms

Journal:	*Transactions on Medical Imaging*
Manuscript ID:	Draft
Manuscript Type:	Full Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Chen, Feiyu; Chongqing University, College of Computer Science Bakic, Predrag; University of Pennsylvania, Radiology Maidment, Andrew; Hospital of the University of Pennsylvania, Dept. of Radiology Jensen, Shane; University of Pennsylvania, The Wharton School Shi, Xiquan; Delaware State University, Computer and Information Sciences Department Pokrajac, David; Delaware State University, Computer and Information Sciences
Keywords:	X-ray imaging and computed tomography < Imaging modalities, Breast < Object of interest, Validation < General methodology
Specialty/Area of Expertise:	anthropomorphic breast phantom, virtual clinical trials, partial volume , Monte Carlo

1 **Description and Characterization of a Novel Method for Partial Volume Simulation in**
2 **Software Breast Phantoms**

3 Feiyu Chen[*], Predrag Bakic[**], Andrew Maidment[**], Shane T. Jensen[***], Xiquan Shi[****], David
4 Pokrajac[****]

5 [*]*College of Computer Science, Chongqing University, Chongqing, China*
6 [**]*Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104*
7 [***]*Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA*
8 *19104*
9 [****]*Computer and Information Sciences Department/Applied Mathematics Research Center,*
10 *Delaware State University, Dover, DE 19901*

11 **NOMENCLATURE**
12 a :=distance of the simulated nipple point from the chest wall.
13 A_i $i=1,2$:= surface areas of the boundary of V_i belonging to planes π_1 and π_2.
14 b := half of the uncompressed phantom thickness.
15 c':=vertical phantom dimension measured above the nipple level.
16 c":= vertical phantom dimension measured below the nipple level.
17 C_i, $i=1, \ldots, K$:= simulated tissue compartments.
18 d := thickness of skin.
19 d_i $i=1,2$:= the distance between a vertex and planes.
20 D := thickness of the simulated Cooper's ligaments.
21 $f_i(\mathbf{x})$, $i=1, \ldots, K$:= compartment shape functions.
22 $F_{ij}(\mathbf{x})$:=difference of the compartment shape functions $f_i(\mathbf{x})$ and $f_j(\mathbf{x})$.
23 $f_M(\mathbf{x})$:= shape function defining the outer surface of the simulated skin layer.
24 $f_m(\mathbf{x})$:= shape function defining the inner surface of the simulated skin layer.
25 MSE_{MCR} := the mean square error estimation of Monte Carlo method using repetition.
26 MSE_{MCS}:= the mean square error estimation of Monte Carlo method based on sample means.
27 MSE_q:=the mean square quantization error.
28 *nPoint*:= the number of intersections between plane and edge(ext).
29 *nVertex, nVertex1, nVertex2*:= the number of vertices above plane.
30 *nVolume* := Number of geometrical shapes need to be computed.
31 $\mathbf{N}$, $\mathbf{N}_1$,$\mathbf{N}_2$:= The normal vectors of approximated planes.
32 N_{MC} := Number of generated random points of Monte Carlo approach within a voxel.
33 N_{MCI} := the number of points that are inside the measured volume.
34 N_{repeat} := the number of repetitions of the Monte Carlo method applied on each voxel.
35 p_0 ,p_1 ,p_2 ,p_i := The percentages of different material in the voxel.
36 $\mathbf{P}_i$, $i=1,\ldots, 8$:= Vertices of Voxel V.
37 $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{S}$, $\mathbf{T}$, $\mathbf{U}$, $\mathbf{W}$, $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$:= Intersections between approximation plane and voxel edges.
38 PV := The sub-volume of voxel V above plane/planes.
39 PV_j , $j=1, \ldots,$ T:= the true value of the partial volume in voxel j.
40 $PV_{A,j}$, $j=1, \ldots,$ T:= the linear approximation of the partial volume in voxel j.
41 $PV_{M,j}$, $j=1, \ldots,$ T := Approximation of the partial volume in voxel j using certain method M.
42 $PV_{MC,j}$, $j=1, \ldots,$ T:= Approximation of the partial volume in voxel j using Monte Carlo.
43 $PV_{MCR,j}$, $j=1, \ldots,$ T:= Estimation of PV_j using Monte Carlo repetition in voxel j.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

44 $PV_{MCR,ij}$, $i=1,\dots N_{repeat}$; $j=1, \dots, T$:= Approximation of the i-th repetition of the Monte Carlo

45 method applied on j-th voxel.

46 q := number of bits to discretize percentage of a partial volume.

47 r_0, r_s := parameters related to compartment orientation and size.

48 s := subsampling factor for naïve reference method; also, a parameter of intersections.

49 S_i, $i=1,2,3$:= surface areas of the boundary formed by the voxel sides.

50 t := The parameter of intersections.

51 T := the total number of partial volume voxels.

52 V := The symbol for a 3D voxel.

53 $|V|$:= The volume of voxel V.

54 $|V_i|$:= The subvolume of material i in the voxel V.

55 $\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2$ = The fixed points on approximating planes.

56 $\mathbf{x}_c$:= The center of voxel V.

57 $\mathbf{x}_m$, $\mathbf{x}_M$:= Two vertices of the voxel, such that one of them inside the skin and another one

58 outside of skin, (used to compute $\mathbf{x}_0$).

59 Δx := linear dimension of voxel V.

60 $\Delta x'$:= linear dimension of voxel V for naïve reference method.

61 ε := difference between partial volumes computed by M and by linear approximation.

62 ε_A:= the error of linear approximation.

63 ε_M := the error of method M.

64 ε_{MC} := the error of Monte Carlo.

65 $\varepsilon_{MCR,ij}$:= the estimate of the error of i-th repetition of the Monte Carlo approach on j-th voxel.

66 μ_V := The X-ray attenuation in the voxel V.

67 μ_i := The X-ray attenuation of material i in the voxel V.

68 $\boldsymbol{\pi}$, $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$:= The linear approximations of boundarie between different materials.

69 $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, $\boldsymbol{\sigma}_3$:= Planes corresponding to voxel sides.

70

71 **Abstract**

72 A modification to our previous simulation of breast anatomy is proposed to improve the quality

73 of simulated x-ray projections images. The image quality is affected by the voxel size of the

74 simulation. Large voxels can cause notable spatial quantization artifacts; small voxels extend the

75 generation time and increase the memory requirements. An improvement in image quality is

76 achievable without reducing voxel size by the simulation of partial volume averaging in which

77 voxels containing more than one simulated tissue type are allowed. The linear x-ray attenuation

78 coefficient of voxels is, thus, the sum of the linear attenuation coefficients weighted by the voxel

79 subvolume occupied by each tissue type. A local planar approximation of the boundary surface

80 is employed. In the two-material case, the partial volume in each voxel is computed by

81 decomposition into up to four simple geometric shapes. In the three-material case, by

82 application of the Gauss-Ostrogradsky theorem, the 3D partial volume problem is converted into

83 one of a few simpler 2D surface area problems. We illustrate the benefits of the proposed

84 methodology on simulated x-ray projections. An efficient encoding scheme is proposed for the

85 type and proportion of simulated tissues in each voxel. Monte Carlo simulation was used to

86 evaluate the quantitative error of our approximation algorithms.

87 **Keywords**: Digital mammography, anthropomorphic breast phantom, partial volume simulation,

88 Monte Carlo.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

I. Introduction

90 This study is motivated by a desire to improve the quality of synthetic images generated using
91 software anthropomorphic breast phantoms. Software breast phantoms have received increasing
92 attention for their use in preclinical validation of breast imaging systems and image analysis
93 methods. Preclinical validation in the form of virtual clinical trials can improve the validation
94 efficacy by identifying the most promising parameter settings to be assessed in a focused clinical
95 trial. There are various designs of software breast phantoms, including phantoms developed
96 using the rules for simulating anatomical structures [1-11] and phantoms based upon individual
97 clinical 3D breast images [12-17].

99 The software anthropomorphic phantoms developed at the University of Pennsylvania have been
100 used in various applications, including the validation and optimization of digital breast
101 tomosynthesis (DBT) reconstruction methods [18-20], DBT image denoising methods [21, 22],
102 ultrasound tomography (UST) reconstruction and segmentation methods [23, 24], analysis of
103 power spectra descriptors in simulated phantom DBT images [10, 25, 26], analysis of texture
104 properties in phantom digital mammography (DM) and DBT images [27, 28], and analysis of
105 tumor detectability in DBT [29-31]. Physical versions of the 3D anthropomorphic software
106 phantom have also been produced [32-36].

108 The current method for simulating breast anatomy [11] assumes that each voxel contains a single
109 tissue type; this may cause notable artifacts due to abrupt attenuation transitions at the borders
110 between regions of different simulated materials. The realism of the resulting phantom images is
111 thus reduced. The realism can be improved by using a smaller voxel size. Reducing the voxel
112 size, however, extends the phantom generation time and increases memory requirements. It
113 should be possible to improve image quality without reducing voxel size by explicitly accounting
114 for voxels containing more than one simulated tissue type.

116 Partial volume (PV) averaging can help reduce the quantization artifacts on boundaries of
117 regions with different simulated materials. In PV averaging, voxels containing more than one
118 simulated tissue type are allowed; thus, the linear x-ray attenuation coefficient of voxels is the
119 sum of the linear attenuation coefficients weighted by the voxel subvolume occupied by each
120 tissue type. The software phantoms in this study have been generated based upon the recursive
121 partitioning of the phantom volume using octrees [11]. Previously, we reported about the
122 development of a PV technique for selected tissue boundaries in our software breast phantoms
123 [37, 38]. In our 2012 SPIE paper, PV simulation was introduced in phantom voxels containing
124 up to two different simulated tissue types [37]. First, the PV of each voxel occupied by different
125 materials was computed, and the linear attenuation coefficient values assigned as the linear
126 combination of attenuations weighted by the PV occupied by each material in the voxel. These
127 PVs could be also used to calculate the proportion of different materials accurately, both in
128 individual voxels and the whole phantom. The same report discussed an encoding technique to
129 accomplish efficient storage of the material composition. The initial results were illustrated
130 using synthetic projections through phantoms with PV simulated on the skin-air boundary only.

132 In our 2012 IWDM paper [38], we proposed an extension to the PV simulation method to
133 include voxels with three simulated materials. The PV was computed based upon a planar
134 boundary approximation in voxels with multiple simulated tissue types. The improvement of

135 image quality was qualitatively validated. The results were shown in the form of slices and
136 simulated X-ray projections of phantoms with and without PV, assuming a parallel beam of
137 monoenergetic x-rays without scatter.
138

139 Our current work is focused upon PV simulation of software phantoms generated based upon
140 rules for simulating anatomical structures [7, 9, 11, 39]. PV simulation has been implicitly used
141 to generate phantoms based upon computed tomography (CT) images of mastectomy specimen
142 [13, 15, 17] or clinical breast CT data [14, 40]. These PV simulations arise naturally, because all
143 raw volumetric images include partial volumes, as various tissues may contribute to the signal
144 acquired in a single image voxel. In the simulation based upon mastectomy CT data [13, 15,
145 17], the values of each reconstructed breast CT image voxel were scaled into a value from 0.01
146 to 0.99. The scaled values were interpreted as percentage of adipose tissue contained in the
147 voxel. The scaling method resulted in phantom images more similar to the original CT data, as
148 compared to the method based upon the segmentation into discrete tissue types. The scaling
149 helped to preserve some of the fine tissue structure which would be lost when using the
150 segmentation; however, it resulted in noisier images. The software phantoms developed using
151 clinical CT data [14, 40] were designed by initially segmenting the CT data into voxels
152 corresponding to skin, adipose tissue, and fibroglandular tissue. To improve realism, it was
153 found to be necessary to segment the fibroglandular tissue into multiple classes based upon CT
154 image intensity level; these classes were associated with different adipose-to-dense tissue
155 volumetric ratios.
156

157 In this paper, we formulate the details of a PV simulation in the general case with up to three
158 tissue types simulated in a voxel. A qualitative validation of the proposed method is performed
159 in the slices through phantoms with PV simulated at different tissue interfaces. In addition, a
160 direct validation is provided by the analysis of the difference between the PV estimates obtained
161 with the proposed method vs. Monte Carlo estimates of PV. Finally, we present results from a
162 qualitative analysis of phantom projections simulated using a polyenergetic divergent x-ray beam
163 approximation without scatter.

164

II. Partial Volume Simulation Method

II.A. Breast Phantom Generation

167 Breast phantoms in this work were generated utilizing the approach described in [11]. The
168 simulated anatomy consists of compartments C_i $i=1,…,K$ and Cooper's ligaments L, which
169 separate the compartments from each other. The distribution, orientation and shapes of the
170 compartments as well as the shape of the Cooper's ligaments are determined by pre-specified
171 shape functions f_i, $i=1,…,K$. The proposed approach utilizes octrees to split the phantom
172 volume V recursively. The recursive partitioning procedure begins with the root node, which is
173 always flagged for splitting. For each level of the tree, we generate the nodes at the next level by
174 recursively splitting the nodes flagged for splitting. The flagged nodes contain more than one
175 material type. The recursive partitioning procedure continues until an individual node of the tree
176 belongs to a single type (non-partial volume nodes) or until the maximal tree level is reached.
177 The nodes corresponding to the maximal tree level correspond to the voxels. The breast outline

178 and skin boundary are simulated with ellipsoidal surfaces, corresponding to the phantom volume
179 vertically above and below the nipple level. The number of compartments K, the shape
180 functions, the skin thickness d and target thickness D of the Cooper's ligaments, and the voxel
181 size Δx are input parameters of the algorithm.

182

183 **II.B. Different cases of phantom voxels containing multiple materials**

184 For realistic cases where $\Delta x < D/\sqrt{3}, \Delta x < d/\sqrt{3}$, the phantom voxels can be categorized as
185 follows (see Fig. 1(a)):

186　　A. Voxels, containing a single material: (1) skin; (2) air; (3) Cooper's ligament; (4) adipose
187　　　　tissue; and (5) fibroglandular dense tissue.
188　　B. Partial volume voxels:
189　　　　a. *Voxels containing two materials (with one bounding surface)*: (6) skin and air;
190　　　　　(7) skin and adipose tissue; (8) skin and dense tissue; (9) skin and Cooper's
191　　　　　ligament; (10) Cooper's ligament and adipose tissue; (11) ligament and
192　　　　　fibroglandular dense tissue
193　　　　b. *Voxels containing three materials (with two bounding surfaces)*: (12) skin,
194　　　　　ligament, and adipose tissue; and (13) skin, ligament, and dense tissue.

195

196

197　　　　　　　　　　　　(a)　　　　　　　　　　　　　　　　　(b)
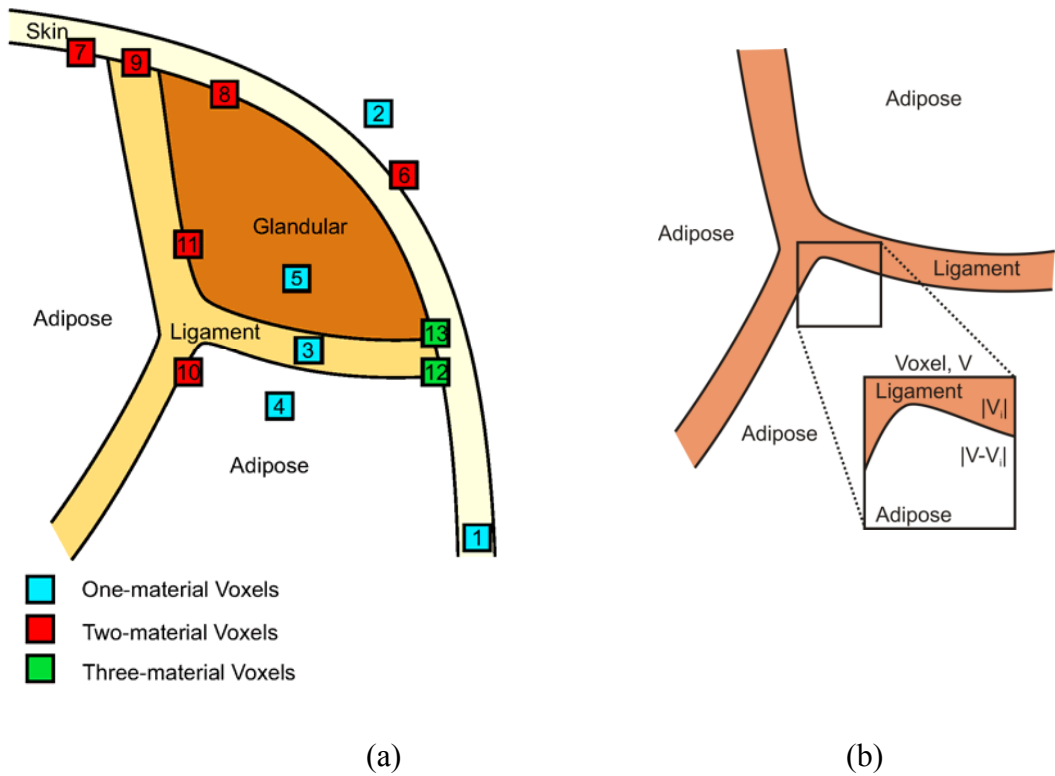
198 **Figure 1:** (a) Taxonomy of material combinations in a voxel. (b) The concept of PV simulation;
199 V denotes the voxel volume and V_i is the sub-volume occupied by dense tissue.
200

201

202 The effective linear x-ray attenuation in a voxel which contains more than one simulated
203 material, Fig. 1(b), can be calculated as:

204
$$\mu_V = \frac{1}{|V|}\sum_i \mu_i |V_i| = \sum_i \mu_i p_i ; \quad p_i = \frac{|V_i|}{|V|} \times 100\% ,$$
(1)

205 where $|V|$ is the voxel volume, $|V_i|$ is the subvolume of material i with the linear x-ray attenuation
206 μ_i, and p_i is the percentage of the material i in the voxel.

207

208 For efficient storage of the voxel material composition, we propose a representation of material
209 types and percentages of the materials using a two-byte binary word. Since a voxel size smaller
210 than the thickness of the skin or Cooper's ligaments is assumed, it is sufficient to consider
211 combinations of up to three materials in a voxel. Thus, it suffices to store percentages of two
212 materials p_1 and p_2. The percentage p_0 of the other material can be calculated by subtracting the
213 stored percentages from 100%, *i.e.*,
214
$$p_0 = 100 - p_1 - p_2.$$
(2)
215 The interpretation of percentages p_0, p_1, and p_2 is specified by a four-bit voxel label, see Table 1.
216 The percentages p_1 and p_2 are stored as two records, q=6 bits each. The choice of q is supported
217 in our results (below); other values of q could be used with this schema as necessary. Using this
218 representation schema, it is possible to encode all partial volume cases from the taxonomy
219 discussed above, see Table 2. For example, consider the case when the label=0 (skin/air
220 boundary). Here, p_1 corresponds to Cooper's ligament tissue (with a constant value 0 in air/skin
221 voxels), while p_2 corresponds to air (the ratio $|V-V_i|/|V|$). The percentage p_0 of skin, can be
222 calculated from Eq. (2). The proposed representation also covers the cases when a voxel is
223 comprised of a single material (e.g., a voxel belonging entirely to skin would have label 0 and
224 $p_1 = p_2 = 0$).

225

226 **Table 1**: Encoding partial volume material percentages using four-bit label

Label		p_0	p_1	p_2
0		skin	ligament	air
1		ligament	fat	dense
2		fat	ligament	skin
3		dense	ligament	skin

227
228

229 **Table 2**: Encoding taxonomy of voxels

Case	p_1 (6 bits)	p_2 (6 bits)	Label (4 bits)
1. Skin	0	0	0
2. Air	0	p_{Air} (=100)	0
3. Cooper's ligament	0	0	1
4. Fat	0	0	2

5. Dense	0	0	3
6. Skin; air	0	p_{Air}	0
7. Skin; fat tissue	0	p_{Skin}	2
8. Skin; dense tissue	0	p_{Skin}	3
9. Skin; Cooper's ligament	p_{Cooper}	0	0
10. Cooper's ligament; fat	p_{Fat}	0	1
11. Cooper's ligament; dense	0	p_{Dense}	1
12. Skin, Cooper's ligament and fat tissue	p_{Cooper}	p_{Skin}	2
13. Skin, Cooper's ligament and dense tissue	p_{Cooper}	p_{Skin}	3

230

231 The rest of this section discusses use of linear approximation to compute partial volumes p_i in
232 two- and three-material partial volume voxels.

233

234 **II.C. Partial volume computation for two material voxels**

235 For voxels containing two materials, we compute a planar approximation of the boundary surface
236 separating the materials. Subsequently, we calculate the portions of the voxel volume split by
237 the planar approximation. Here we discuss the planar approximations for voxels containing skin
238 (Cases 6-9) and voxels on ligament-compartment boundaries (Cases 10-11), followed by the
239 computation of the voxel's volume above the plane.

240

241 *II.C.1. Voxels containing skin (Cases 6-9; see Table 2)*

242 We assume that the outer and the inner surface of the skin (skin/air and skin/tissue boundaries),
243 are defined by functions $f_M(\mathbf{x})$ and $f_m(\mathbf{x})$ as follows, respectively [11].

244
$$f_m(\mathbf{x}) = f_m(x,y,z) = \begin{cases} 1, x < 0 \\ \dfrac{x^2}{(a-d)^2} + \dfrac{y^2}{(b-d)^2} + \dfrac{z^2}{(c'-d)^2}, x > 0, z \ge 0, \\ \dfrac{x^2}{(a-d)^2} + \dfrac{y^2}{(b-d)^2} + \dfrac{z^2}{(c''-d)^2}, x > 0, z < 0 \end{cases} \tag{3}$$

245
$$f_M(\mathbf{x}) = f_M(x,y,z) = \begin{cases} 1, x < 0 \\ \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c'^2}, x > 0, z \ge 0, \\ \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c''^2}, x > 0, z < 0. \end{cases} \tag{4}$$

246   Here we discuss in detail the computation of the partial volume for voxels containing skin and air

247   (case 6). Other cases can be treated similarly (with function $f_M(\mathbf{x})$ appropriately replaced with

248   $f_m(\mathbf{x})$). Since $f_M(\mathbf{x})$ is known in a closed form, the volume $|V_i|$ can be exactly calculated;

249   however, this calculation is computationally inefficient. Instead, the function $f_M(\mathbf{x})$ is

250   approximated by a tangent plane $\boldsymbol{\pi}$ which reduces the considered problem to computation of the

251   voxel volume below a pre-specified plane. The tangent plane $\boldsymbol{\pi}$ on the surface $f_M(\mathbf{x})=1$ is placed

252   at a point $\mathbf{x}_0$ inside the voxel V. The point $\mathbf{x}_0$ satisfies $f_M(\mathbf{x})=1$ and is on the line segment between

253   the points $\mathbf{x}_m$ and $\mathbf{x}_M$. The points $\mathbf{x}_m$ and $\mathbf{x}_M$ are calculated such that $\mathbf{x}_m=\arg \min_V f_M(\mathbf{x})$,

254   $\mathbf{x}_M=\arg \max_V f_M(\mathbf{x})$. Points $\mathbf{x}$ on the tangent plane $\boldsymbol{\pi}$ satisfy: $(\mathbf{x} - \mathbf{x}_0) \cdot \mathbf{N} = 0$, where $\cdot$ denotes

255   the scalar product and $\mathbf{N} = \nabla f_M(\mathbf{x}_0)$ is the gradient vector at the point $\mathbf{x}_0$ (See Fig. 2)
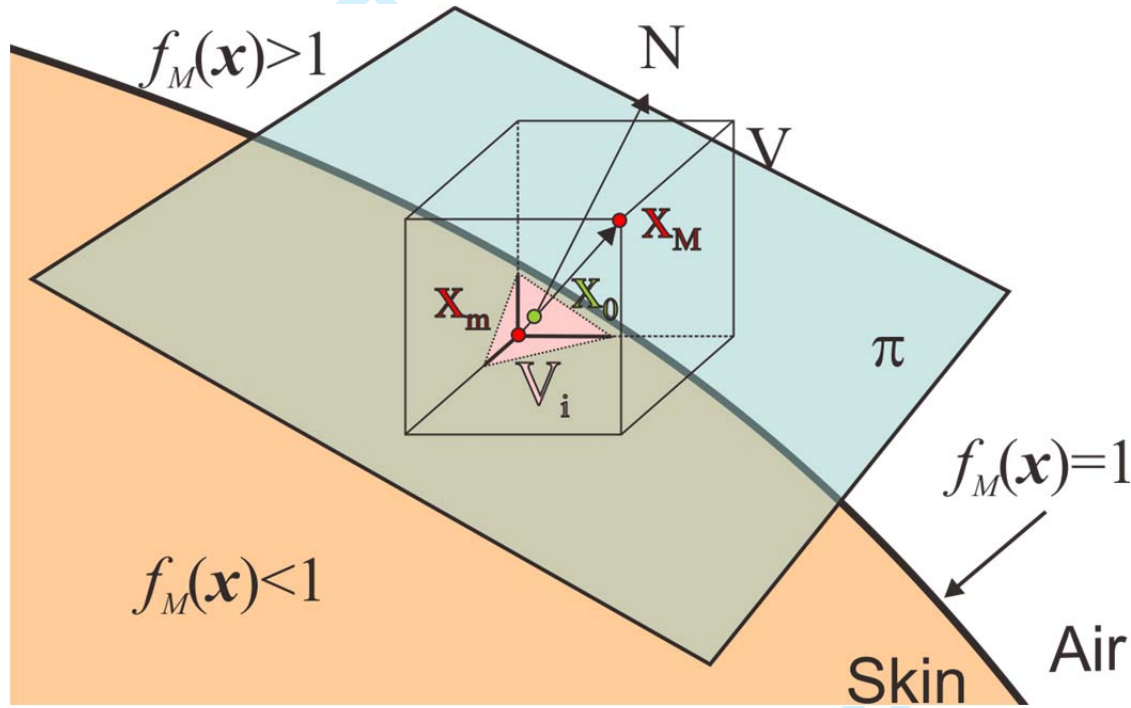
256   **Figure 2**: Local approximation of skin boundary (defined by $f_M(\mathbf{x}) = 1$) by a tangent plane.

257

II.C.2. Voxels containing Cooper's ligaments and compartmental tissue (Cases 10-11; see Table 2)

260   The planar approximation for the boundary between the Cooper's ligaments and adipose tissue

261   or dense tissue (cases 11 and 12) can be obtained as follows. Without loss of generality,

262   consider adipose compartments C_i and C_j with corresponding shape functions f_i and f_j. A

263   Cooper's ligament between the compartments is the locus of points within a distance D/2 from a

264   median surface $F_{ij}(\mathbf{x})= f_i(\mathbf{x})- f_j(\mathbf{x})=0$, see Fig. 3. Consider a voxel V with center $\mathbf{x}_c$. We define a

265   planar approximation $\boldsymbol{\pi_1}$ of the boundary between the Cooper's ligament and the compartment

266   C_j as

267 $$\pi_1 : \left(\mathbf{x} - \mathbf{x}_1 \right) \cdot \mathbf{N}_1 = 0 \, . \tag{5}$$

268 Here, $\mathbf{N}_1$ is a vector normal to a level set $F_{ij}(\mathbf{x}) = F_{ij}(\mathbf{x_c})$ at x_c. x_1 is a point on the normal at
269 distance D from the intersection of the normal and the median surface, such that:

270 $$\mathbf{x}_1 = \mathbf{x}_c + sign\left(F_{ij}\left(\mathbf{x}_c\right)\right) \left(D/2 - \frac{F_{ij}\left(\mathbf{x}_c\right)}{\left\|\nabla F_{ij}\left(\mathbf{x}_c\right)\right\|} \right) \cdot \frac{\nabla F_{ij}\left(\mathbf{x}_c\right)}{\left\|\nabla F_{ij}\left(\mathbf{x}_c\right)\right\|}, \tag{6}$$

$$\mathbf{N}_1 = sign\left(F_{ij}\left(\mathbf{x}_c\right)\right) \nabla F_{ij}\left(\mathbf{x}_c\right).$$

271

272
273
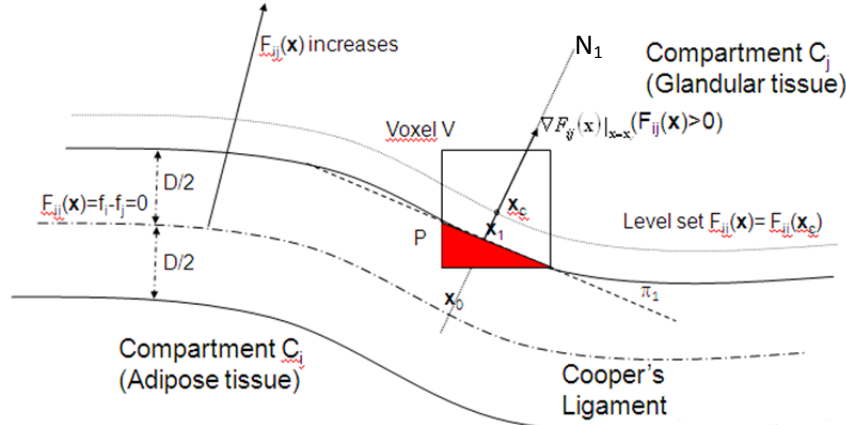
274 **Figure 3:** Planar approximation of a boundary between Cooper's ligament and a compartment.

275

II.C.3 Calculation of the volume above a planar approximation of two materials boundary

276

277 In this section, a fast and exact method is proposed to determine the fraction $|V_i|$ of the voxel's V
278 volume located above the plane π (a case when V_i is below the plane is reduced to this case by
279 changing the direction of the normal vector of the plane). The first step of the method is to
280 determine the number $nVertex$ of voxel vertices above the plane. The voxel vertices $\mathbf{P}_i$
281 above the plane π specified by a normal $\mathbf{N}$ and containing a point x_0 satisfy:

282 $$\left(\mathbf{P}_i - \mathbf{x}_0 \right) \cdot \mathbf{N} > 0 \tag{7}$$

283 Depending upon $nVertex$, $|V_i|$ is computed using fundamental geometric shapes (e.g., prisms,
284 prismoids or tetrahedrons, see Fig. 4). Note that if $nVertex>4$ computation of $|V_i|$ is reduced to
285 computation of the complementary volume to the partial volume of V below the plane, see
286 Algorithm A1 (Appendix 1). The detailed algorithm for computation of partial volume of a
287 voxel V above a given plane is specified in Algorithm A2
288
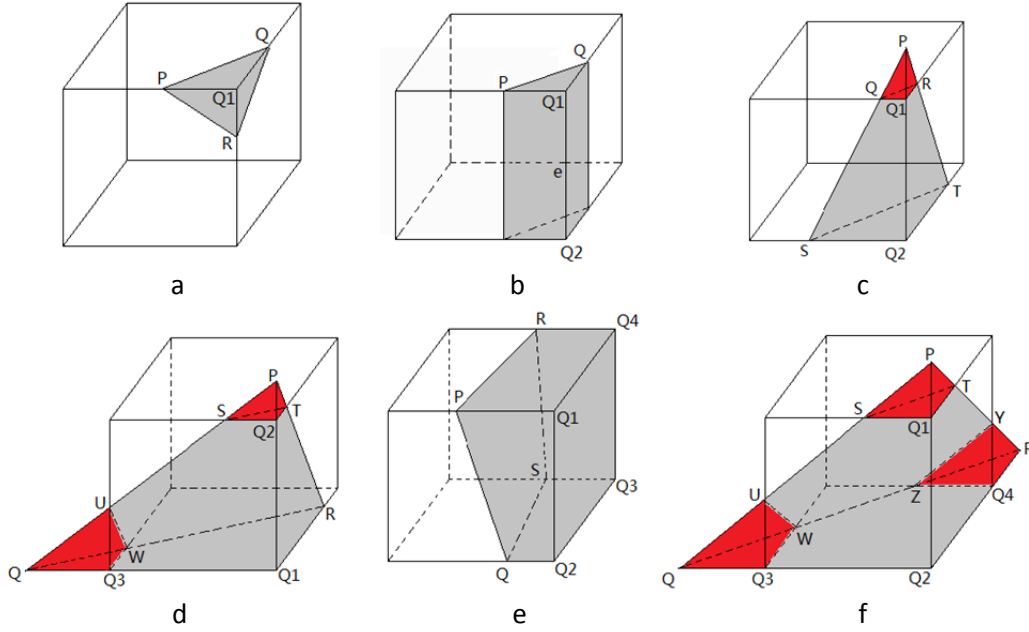
Figure 4: Different cases of sub-volume.

The Algorithm A2 is very efficient. Observe that the considered partial volume problem reduces to 6 cases (Fig. 4). In each case, a small number of intersections (up to 9) between the plane π and voxel's edges (their extensions) need be calculated, see Table 3, followed by computation of a volume of a geometric primitive.

Computation of intersections is also fast. To compute an intersection $\mathbf{Q}$ between an edge (extension of edge) containing vertices $\mathbf{P}_i$ and $\mathbf{P}_j$ and the plane π, it is sufficient to resolve the system:

$$\mathbf{Q} = \mathbf{P}_i + t(\mathbf{P}_i - \mathbf{P}_j)$$
$$(\mathbf{Q} - \mathbf{x}_0) \cdot \mathbf{N} = 0$$

which results in the parameter t specified by:

$$t = -\frac{(\mathbf{P}_i - \mathbf{x}_0) \cdot \mathbf{N}}{(\mathbf{P}_i - \mathbf{P}_j) \cdot \mathbf{N}}. \tag{8}$$

Since the vertices $\mathbf{P}_i$ and $\mathbf{P}_j$ differ in only one coordinate, this requires computation of only one scalar product $((\mathbf{P}_i - \mathbf{x}_0) \cdot \mathbf{N})$.

The value of parameter t depends on the position of $\mathbf{Q}$. If $\mathbf{Q}$ is located between $\mathbf{P}_i$ and $\mathbf{P}_j$, then $0 < t < 1$.

To compute intersection between π and other edges (extensions) we may proceed as follows. For an intersection $\mathbf{R} = \mathbf{P}_i + s(\mathbf{P}_i - \mathbf{P}_j)$ between the plane and edge (extension of edge) containing $\mathbf{P}_j$ and $\mathbf{P}_k$, it is sufficient to compute

1
2
3
4   312
5
6
7
8   313
9   314
10
11   315
12
13   316
14   317
318

15   319
16   320
17
18   321
19   322
20   323
21   324
22
23   325

$$s = -\frac{(\mathbf{P}_k - \mathbf{x}_0) \cdot \mathbf{N}}{(\mathbf{P}_k - \mathbf{P}_j) \cdot \mathbf{N}} = \frac{(\mathbf{P}_i - \mathbf{P}_k) \cdot \mathbf{N} - (\mathbf{P}_i - \mathbf{x}_0) \cdot \mathbf{N}}{(\mathbf{P}_k - \mathbf{P}_j) \cdot \mathbf{N}}$$

which does not require computation of additional scalar products since $\mathbf{P}_i$-$\mathbf{P}_j$ only contains one non-zero coordinate.

The advantage of this procedure is that we can easily compute the partial volume without considering the shape of the boundary (interface) and the number of intersections between the plane and voxel. The cases are distinguished based on $nVertex$, that can be obtained easily.

Table 3: Number of vertices above the planar approximation of the material boundary ($nVertex$), number of intersections between the approximation and the voxel edges or edge extensions ($nPoint$) and number of volumes of geometrical primitives to be computed ($nVolume$), for different cases of Algorithm A2 (see Appendix 1).

	Case A	Case B1	Case B2	Case C	Case D1	Case D2
$nVertex$	1	2		3	4	
$nPoint$	3	4	5	7	4	9
$nVolume$	1(tetrahedron)	1(prism)	2(tetrahedron)	3(tetrahedron)	1(prismoid)	4(tetrahedron)

326

327   **II.D. Partial volume computation for three material voxels (Cases 12, 13; see Table 2)**

328   For a voxel V containing three materials we construct a planar approximation for each bounding
329   surface (Fig. 5). The results of the approximation are planes π_1: $(\mathbf{x} - \mathbf{x}_1) \cdot \mathbf{N}_1 = 0$ and π_2:
330   $(\mathbf{x} - \mathbf{x}_2) \cdot \mathbf{N}_2 = 0$. Here, π_1 and π_2 are linear approximations of the inner skin boundary and the
331   ligament's boundary, respectively. The partial volume $|V_i|$ of interest is subsequently calculated
332   as the volume of a portion of the voxel V that is below/above the planes. For example, the partial
333   volume V_i corresponding to the adipose tissue in Fig. 5 is computed as a volume of a part of the
334   voxel that is both above planes π_1 and π_2.

335
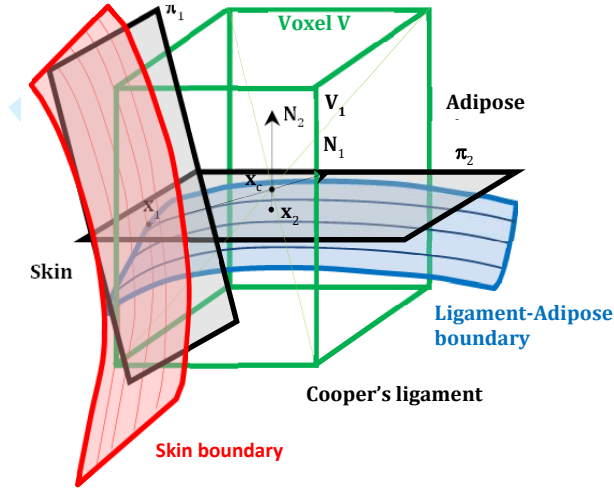
Figure 5: An illustration of a three material voxel containing skin, Cooper's ligament and adipose tissue and planar approximations π_1 and π_2 of the tissue boundaries.

Given the planar approximations π_1 and π_2 of the material boundaries, we compute the partial volume $|V_i|$ using the divergence (i.e., Gauss-Ostrogradsky) theorem [41, 42] . Without loss of generality, we consider the volume V_i that is above both planes π_1 and π_2 (other cases can be treated by changing directions of vectors specifying π_1 and π_2). The divergence theorem can be stated as the following integral equation:

$$\iiint_{V_i}(\nabla \cdot F)dV = \iint_S (F \cdot \mathbf{N})dS . \tag{9}$$

The left side is a volume integral of a vector field $F(\mathbf{x})$ over the partial volume V_i, the right side is the surface integral over the boundary of the volume V_i, and $\mathbf{N}$ is the outward pointing unit normal vector of the boundary. Note that the volume V_i is bounded by planes π_1 and π_2 and at most 6 sides of the voxel.

The application of the divergence theorem depends on whether there is a voxel vertex $\mathbf{Q}_1$ above both planes π_1 and π_2. Assume that such a vertex $\mathbf{Q}_1$ exists. By choosing $F(\mathbf{x}) = \mathbf{x}$, Eq. (9) reduces to:

$$3|V_i| = (S_1 + S_2 + S_3)\Delta x + A_{\pi 1}d_1 + A_{\pi 2}d_2 \tag{10}$$

where S_i, $i=1,\ldots,3$ are surface areas of the boundary formed by the voxel sides σ_1 ,σ_2 and σ_3 that do not contain the vertex $\mathbf{Q}_1$; and A_1 and A_2 are surface areas of the boundary of V_i belonging to planes π_1 and π_2 and $d_1 = (\mathbf{Q}_1 - \mathbf{x}_1) \cdot \mathbf{N}_1$, and $d_2 = (\mathbf{Q}_1 - \mathbf{x}_2) \cdot \mathbf{N}_2$ are distances of the vertex $\mathbf{Q}_1$ to planes π_1 and π_2 (see Fig. 6). Subsequently, the PV is calculated as:

$$|V_i| = \frac{(S_1 + S_2 + S_3)\Delta x + A_{\pi 1}d_1 + A_{\pi 2}d_2}{3} . \tag{11}$$

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32   357
33
34   358
35   359
36   360
37
38   361
39   362
40   363
41   364
42
43   365
44   366
45
46
47
48
49
50
51
52
53
54
55
56
57
58   367
59
60

Figure 6: Partial volume V_i of the voxel V above planes π_1 and π_2 and containing vertex $\mathbf{Q}_1$. S_1, S_2 and S_3 (here $S_3=0$) are surface areas of parts of the volume boundary belonging to voxel sides σ_1, σ_2 and σ_3 that do not contain the vertex $\mathbf{Q}_1$.

If there is no vertex $\mathbf{Q}_1$ of V above both planes π_1 and π_2, it is still possible that $|V_i|>0$. As illustrated in Fig. 7, this is the case when the sets of vertices above π_1 and π_2 are both non empty. In such a case, partial volume can be computed as the difference of partial volumes above one of the planes (e.g., plane π_1), calculated using Algorithm A1) the partial volume above π_1 and *below* π_2 (calculated by changing the direction of the normal vector $\mathbf{N}_2$).
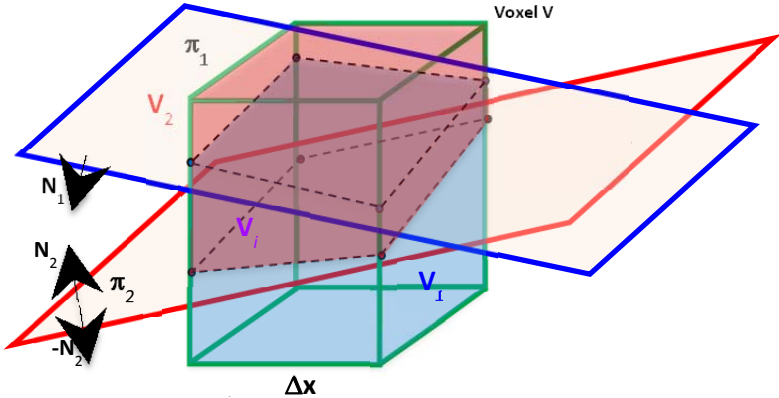
368 **Figure 7:** Illustration of a case when there is no vertex of a voxel V above both planes π_1 and π_2
369 but the partial volume is larger than zero. The volume V_i is the intersection of V_1 and V_2.
370
371 The Algorithm A3 (Appendix 1) for computation of voxel partial volume above two planes is
372 also very efficient. The 3D partial volume problem is converted into the computation of the
373 linear combinations of a few 2D polygon areas. The polygon's vertices are chosen from the
374 vertices of the voxel; the intersections of an edge of voxel and the plane; or the intersections of
375 two planes on a voxel side $\sigma_i, i = 1,2,3$. The intersection of an edge of the voxel and the plane
376 can be solved using Eq. (8).

377 **III. VALIDATION TECHNIQUES**

378 We have performed two validations of the proposed technique. First, we tested to what extent
379 the proposed technique of linear approximation was capable of adequately representing the true
380 values of partial volumes. This was performed by quantitative validation of the algorithm.
381 Second, we evaluated the improvement of image quality by visual assessment of simulated
382 images of phantoms with PV.

383 **III.A. Accuracy assessment of the PV computation**
384
385 The goal of qualitative validation of the proposed algorithm for PV approximation is to estimate
386 the expectation of the squared error, i.e., $E(\varepsilon_A{}^2)$, where, for each partial volume voxel j, the
387 estimation error is defined as:

388 $$\varepsilon_{A,j} = PV_j - PV_{A,j}. \tag{12}$$

389 Here $PV_{A,j}$ is the partial volume in voxel j obtained using the proposed method (a linear
390 approximation) and PV_j is the true value of the partial volume, where (see Section II) both PV_j
391 and $PV_{A,j}$ belong to a range [0, 1]. However, a practical problem is that the true partial volumes
392 PV_j are not directly observable, hence $\varepsilon_{A,j}$ cannot be directly computed.

393 Consider a reference method M that can estimate the partial volume $PV_{M,j}$ for each voxel j. At
394 voxel j, we can only directly observe the difference ε_j between partial volumes computed by
395 method M and by the linear approximation:

$$\varepsilon_j = PV_{M,j} - PV_{A,j}.$$

397 We can easily obtain that:

398 $$\varepsilon_j = \varepsilon_{M,j} + \varepsilon_{A,j} \tag{13}$$

399 where:

400 $$\varepsilon_{M,j} = PV_{M,j} - PV_j \tag{14}$$

401 From Eq. (13):

402 $$E(\varepsilon_A{}^2) = E(\varepsilon^2) - E(\varepsilon_M{}^2) - 2\,E(\varepsilon_M\,\varepsilon_A). \tag{15}$$

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
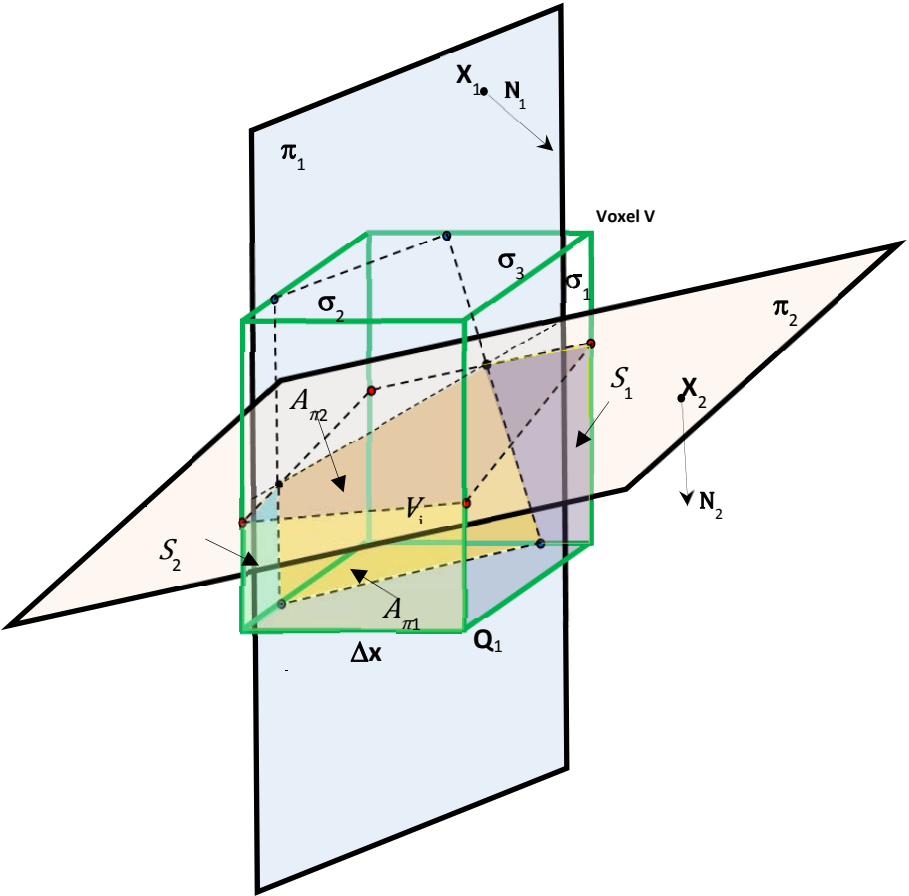41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

403 Note that the definition of $E(\varepsilon_A{}^2)$ guarantees that both sides of Eq. (15) are always non-negative.
404 Also, note that the expectations are calculated over one particular phantom. In order to use a
405 reference method M to estimate $E(\varepsilon_A{}^2)$ using Eq. (13), one should be able to estimate the squared
406 error of the reference method, $E(\varepsilon_M{}^2)$. Also, the errors of the proposed approximation and the
407 reference method should be uncorrelated, i.e., $E(\varepsilon_M \varepsilon_A)=0$.

408 A naïve choice for the reference method M is estimation of the PV based on subsampling. Let P
409 be a considered partial volume phantom (with linear voxel dimension Δx). Consider a non-
410 partial volume phantom P' that simulates an identical anatomy as P, with linear voxel dimension
411 $\Delta x' = \Delta x/s$ where s is an integer subsampling factor. For each voxel j, the partial volume can be
412 estimated as the fraction of the corresponding voxels from P' that contain the material of interest.
413 Unfortunately, the subsampling method is not suitable as a reference method. First, in this
414 method, $E(\varepsilon_M{}^2)$ cannot be easily estimated. Second, errors of the proposed approximation and
415 the reference method are correlated (e.g., ε_M is larger when the boundaries between different
416 materials are more non-linear, i.e., where ε_A is larger). Finally, the method may not be feasible,
417 since the computation of a phantom P' for large s may be computationally prohibitive.
418
419 To overcome these difficulties, we propose to utilize a Monte Carlo approach [43] which gives
420 us the opportunity to compare the precision of our PV approximation with a reference method
421 based on Monte Carlo simulation. For each partial volume voxel j, we estimate $PV_{MC,j}$ as follows.
422 We generate N_{MC} random points $\mathbf{x}$ within a voxel and determine the number N_{MCI} of points that
423 are inside the measured partial volume. Note that for voxels from cases 6-9 (see section 2.3.1)
424 this includes computing functions $f_M(\mathbf{x})$ ($f_m(\mathbf{x})$) (Eq. (3, 4). For PV voxels containing
425 ligament tissue (cases 10-13) we, in addition, need to determine *the exact* distance between $\mathbf{x}$ and
426 the median surface (see Section 2.3.2); this can be done, e.g., using the algorithm described in
427 [44]. The partial volume is subsequently obtained as

428 $\quad PV_{MC,j}=N_{MCI}/N_{MC}.$ (16)

429 The error of the Monte Carlo method is defined as:

430 $\quad \varepsilon_{MC,j}=PV_{MC,j}-PV_j$ (17)

431 As shown in the Appendix 2, Eq. (A8), $E(\varepsilon_{MC} \varepsilon_A)=0$.

432 Therefore, from Eq. (15) we obtain:

433 $\quad E(\varepsilon_A{}^2)= E(\varepsilon^2)-E(\varepsilon_{MC}{}^2).$ (18)

434 $E(\varepsilon^2)$ can be estimated using the sample mean square error (MSE):

435 $\quad E(\varepsilon^2) \sim MSE_{total}=\frac{1}{T}\sum_{j=1}^{T} \varepsilon_j^2 ,$ (19)

436 where T denotes the total number of partial volume voxels. The following subsection discusses
437 estimates MSE_{MC} of $E(\varepsilon_{MC}{}^2)$. According to Eq. (18), (19) we estimate $E(\varepsilon_A{}^2)$ as:

438 $\quad MSE_A= MSE_{total} -MSE_{MC}.$ (20)

439 Note that the computed partial volumes are quantized using q bits. Hence, $E(\varepsilon_A^2)$ is bounded by
440 the quantization error. Under the assumption that $PV_{M,j}$ are uniformly distributed within each
441 quantization interval [45], the quantization error MSE_q is approximated as:

442
$$MSE_q = \frac{1}{12(2^q-1)^2}.$$
(21)

III.A.1. Estimation of MSE of Monte Carlo approach

443

444 To ensure reliability of the validation, we utilize two techniques to estimate MSE of Monte Carlo
445 approach. The first technique repeats the Monte Carlo process for each voxel in order to assess
446 the true value of ε_{MC}. The second technique is based on estimation of sample means of
447 computed PV_{MC} and completely avoids estimation of ε_{MC}.

448   1) MSE of Monte Carlo based on estimating ε_{MC}

449 Consider voxels belonging to one specific case of partial volume (e.g., two-material voxels on
450 the ligament-compartment boundary). For each partial volume voxel of a particular phantom, we
451 repeat the estimation of PV using the Monte Carlo approach N_{repeat} times. Denote the obtained
452 estimates in the i-th repetition of the Monte Carlo method applied on j-th voxel as $PV_{MCR,ij}$
453 ($i=1,\ldots,N_{repeat}$, $j=1\ldots T$). The idea of this approach is to obtain the estimate of the true partial
454 volume by averaging these estimates.

455 Since the MC estimation is unbiased (see Eq. (A1)), we can use the following estimate of the
456 true partial volume PV_j for the j-th voxel:

457
$$PV_{MCR,j} = \frac{1}{N_{repeat}} \sum_{i=1}^{N_{Repeat}} PV_{MCR,ij}$$
(22)

458 Based on this, the estimate $\varepsilon_{MCR,ij}$ of the error of the i-th repetition of the Monte Carlo approach
459 on j-th voxel is:

460
$$\varepsilon_{MCR,ij} = PV_{MCR,ij} - PV_{MCR,j}$$
(23)

461 Therefore, the MSE of MC can be estimated as:

462
$$MSE_{MCR} = \frac{1}{TN_{repeat}} \sum_{j=1}^{T} \sum_{i=1}^{N_{Repeat}} \varepsilon_{MCR,ij}^2 = \frac{1}{TN_{repeat}} \sum_{j=1}^{T} \sum_{i=1}^{N_{Repeat}} (PV_{MCR,ij} - PV_{MCR,j})^2$$
(24)

463 In practice, it may be more computationally efficient to utilize the following formula:

464
$$MSE_{MCR} = \frac{1}{TN_{repeat}} \sum_{j=1}^{T} [\sum_{i=1}^{N_{Repeat}} PV_{MCR,ij}^2 - \frac{1}{N_{repeat}} (\sum_{i=1}^{N_{Repeat}} PV_{MCR,ij})^2]$$
(25)

465   2) Estimation based on sample means of PV_{MC}

466 Here we demonstrate that an estimate of $E(\varepsilon_{MC}^2)$ can be obtained by estimating $E(PV_{MC})$ and
467 $E(PV_{MC}^2)$. Note that this approach does not require knowledge of true values of PV_j at each
468 voxel nor the availability of distribution $p(PV)$.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

469   From Eq. (17) follows:

470     $$E(PV^2) = E(PV_{MC}^2) - E(\varepsilon_{MC}^2) - 2E(\varepsilon_{MC}PV) \tag{26}$$

471   which due to Eq. (A5) in the Appendix 2 reduces to:

472     $$E(PV^2) = E(PV_{MC}^2) - E(\varepsilon_{MC}^2). \tag{27}$$

473   Using Eq. (A7) and Eq. (A4), from Eq. (27) we obtain:

474     $$E(\varepsilon_{MC}^2) = \frac{1}{N_{MC}-1}\left(E(PV_{MC}) - E(PV_{MC}^2)\right). \tag{28}$$

475   We estimate $E(\varepsilon_{MC}^2)$ using MSE_{MCS} computed by utilizing sample means of Monte Carlo
476   estimations of the partial volume and of the squared partial volume as:

477     $$MSE_{MCS} = \frac{1}{N_{MC}-1}\left(\frac{1}{T}\sum_{j=1}^{T} PV_{MC\,j} - \frac{1}{T}\sum_{j=1}^{T} PV_{MC\,j}^2\right) \tag{29}$$

478   **III.B. Image quality improvement after PV simulation**
479   All the simulations were implemented using *Matlab* (64-bit, MathWorks, Natick, MA).
480   Phantoms were simulated on a computer with two Intel Xeon 5650 Six Core Processors (Intel,
481   Inc., Santa Clara, CA) working at 2.53 GHz with 128 GB RAM (1333 MHz DDR III ECC) and
482   utilizing one core per phantom. We used Matlab version v7.13(R2011b).
483

484   We generated 450ml phantoms (approximately a B cup bra size), [46] with ellipsoidal outline
485   semiaxes a=b=c''=5cm, c'=12cm, (see Eqs. (2) and (3)). The voxel size were 100μm and
486   200μm. The number of compartments was K=333 [11].
487

488   We specified the skin thickness d =1.5mm based upon reports in the literature and the target
489   thickness of the Cooper's ligaments D=0.6mm [47, 48]. There are no explicit quantitative
490   reports in the literature on the measured thickness of Cooper's ligaments in clinical data. We
491   assumed the thickness was smaller than 1 mm, as observed from subgross breast histological
492   sections (e.g., the sections shown in [39]). Also, we varied the relative random compartment
493   orientation r_0 and the relative compartment size, r_s [49].
494

495   Mammographic images of the phantom are simulated using (i) a finite-element model of
496   mammographic breast compression, and (ii) simulation of the x-ray projections through the
497   compressed phantom. The deformation model is implemented using Abaqus (version 6.6, DS
498   Simulia Corp., Providence, RI), and is based upon a finite element model of breast compression
499   proposed by Ruiter *et al.* [50]. The deformation model assumes the volume of the simulated
500   breast tissue is preserved. With that assumption, a 450 ml phantom described in Section II.B
501   corresponds to a compressed phantom with a size of 20 cm in the vertical direction, 5 cm in the
502   lateral direction, and approximately 6.5 cm in the chest wall-nipple direction. Mammographic
503   projections of the compressed phantom are simulated assuming a polyenergetic x-ray acquisition
504   model without scatter. The quantum noise was simulated by a random Poison process,
505   corresponding to the standard radiation dose of a clinical mammographic projection. The linear

506 x-ray attenuation coefficients of the simulated tissues were selected using their energy
507 dependence as listed in the NIST X-ray Mass Attenuation Tables [51]. The simulated
508 acquisition geometry uses a source–detector distance of 70 cm, a detector element size of 70 μm,
509 and a 24 cm × 30 cm field-of-view, corresponding to the Hologic Selenia Dimensions full-field
510 digital mammography system (Hologic, Bedford, MA).

511

IV. RESULTS

IV.A. Qualitative evaluation

514 Fig. 8 illustrates the PV simulation in a 450ml software breast phantom with 200μm voxels,
515 relative compartment orientation, $r_0 \in [0.5,2]$ and relative compartment size, $r_s \in [0.5,2]$.
516 Shown is the segmentation of phantom detail into air and voxels containing one, two or three
517 materials.

518

(a)

(b)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16 523
17
18
19
20 524
21
22
23 525
24 526
25 527
26
27
28 529
29 530
30 531
31 532
32 533
33
34 534
35 535
36 536
37
38 537
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
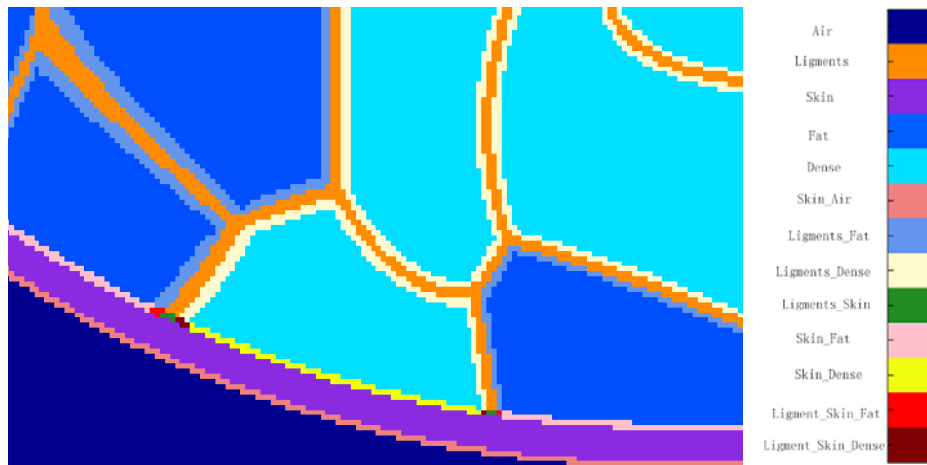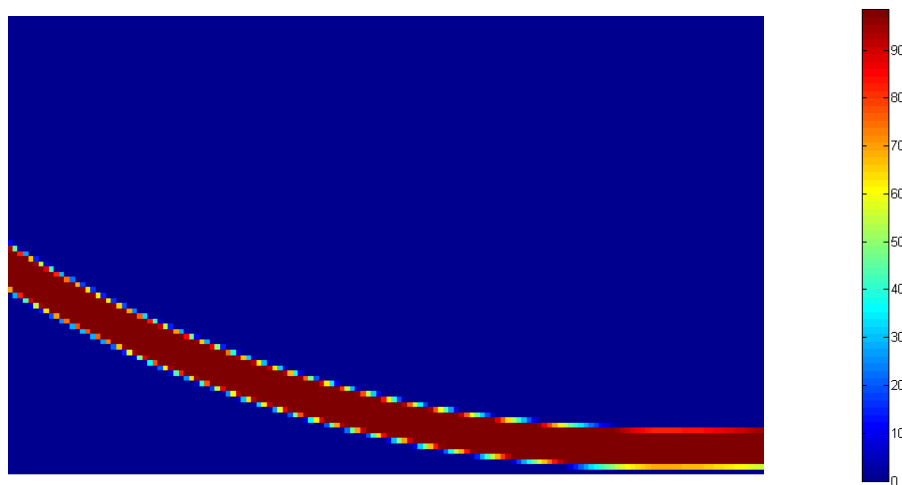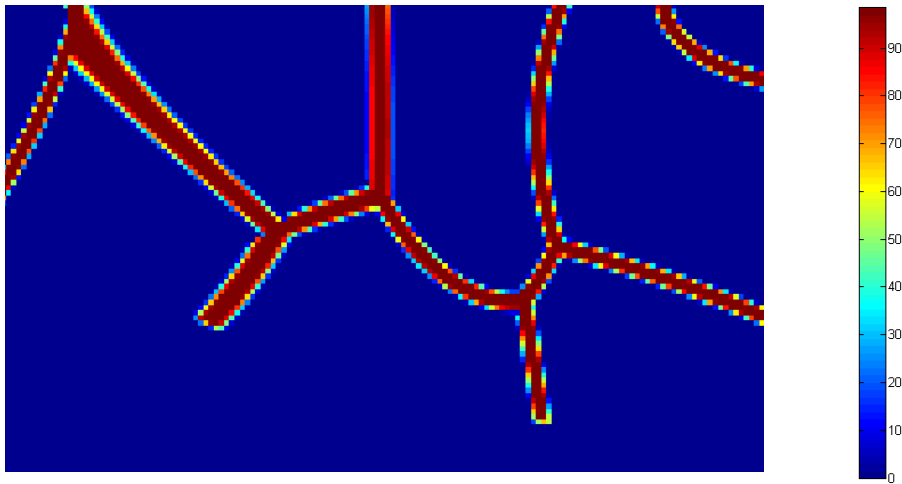
(c)

Figure 8: (a) Various cases of PV voxels simulated by the proposed method in a slice of a 450ml software breast phantom, with 200μm voxels. Color-coded percentage of the skin (b) or Cooper's ligaments (c) have been computed in phantom voxels from (a).

Fig. 9 shows simulated x-ray projections of phantoms with and without simulated PV. The simulated acquisition assumed a polyenergetic x-ray beam and divergent x-ray beam. The projections correspond to three phantoms with identical distributions of compartments: a phantom with 200μm voxels and no PV (Fig. 9(a)); a 200μm phantom with simulated PV (Fig. 9(b)); and a phantom with 100μm voxels and no PV (Fig. 9(c)). Fig. 9(d) contains magnified detail of Fig. 9(a). Corresponding magnified details of Fig. 9(b) and Fig. 9(c) are shown in Fig. 9(e) and Fig. 9(f), respectively. Fig. 10 illustrates the effect of simulating PV with respect to the skin, two-material voxels, and three-material voxels.
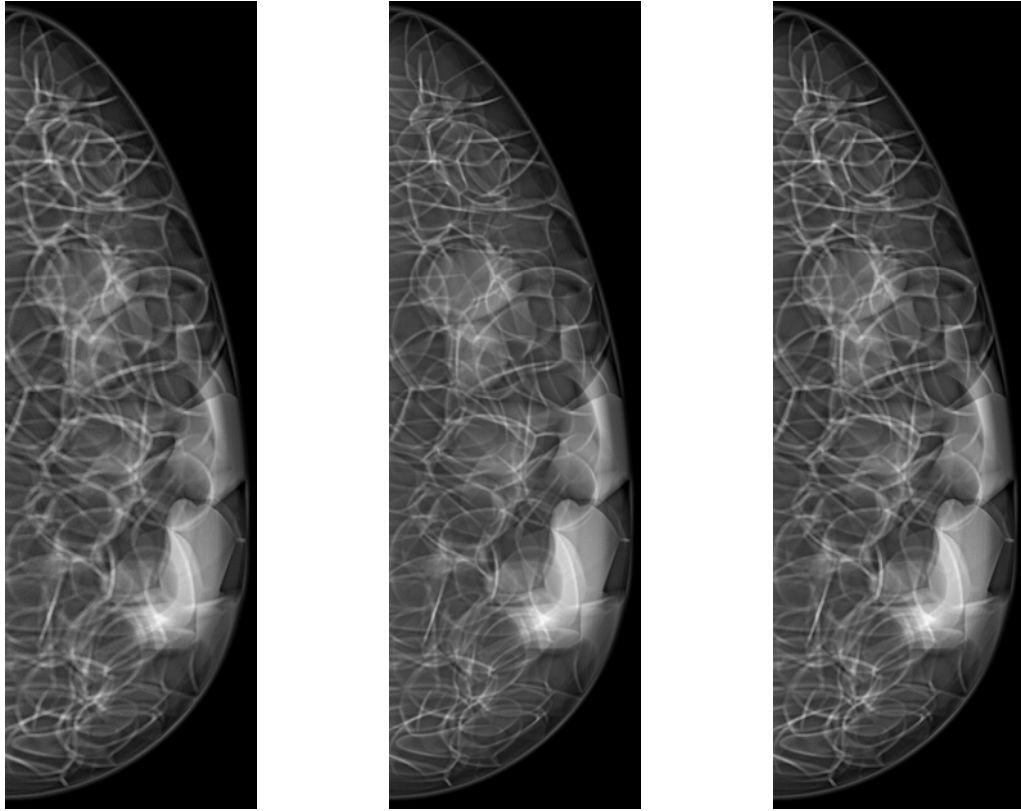
(a)　　　　　　　　　　(b)　　　　　　　　　　(c)
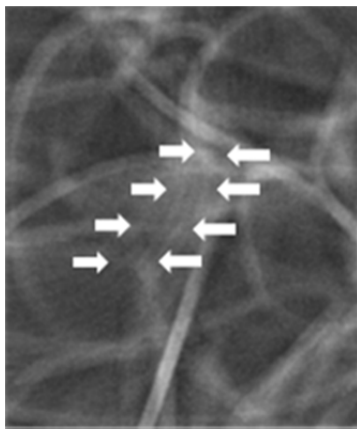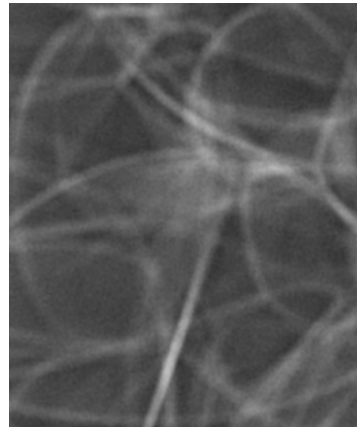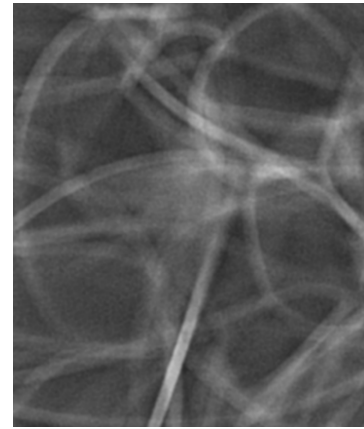
(d)　　　　　　　　　　(e)　　　　　　　　　　(f)

Figure 9: Simulated projections of: (a) a phantom with 200μm voxels and no PV; (b) the phantom from (a) with simulated PV; (c) the same phantom generated at 100 μm voxels and no PV; (d) A magnified detail from Fig. (a) (white arrows indicate stair-step quantization artifacts); (e) the corresponding detail from Fig. (b); and (f) the corresponding detail of Fig. (c).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

546

547

(a)   (b)   (c)
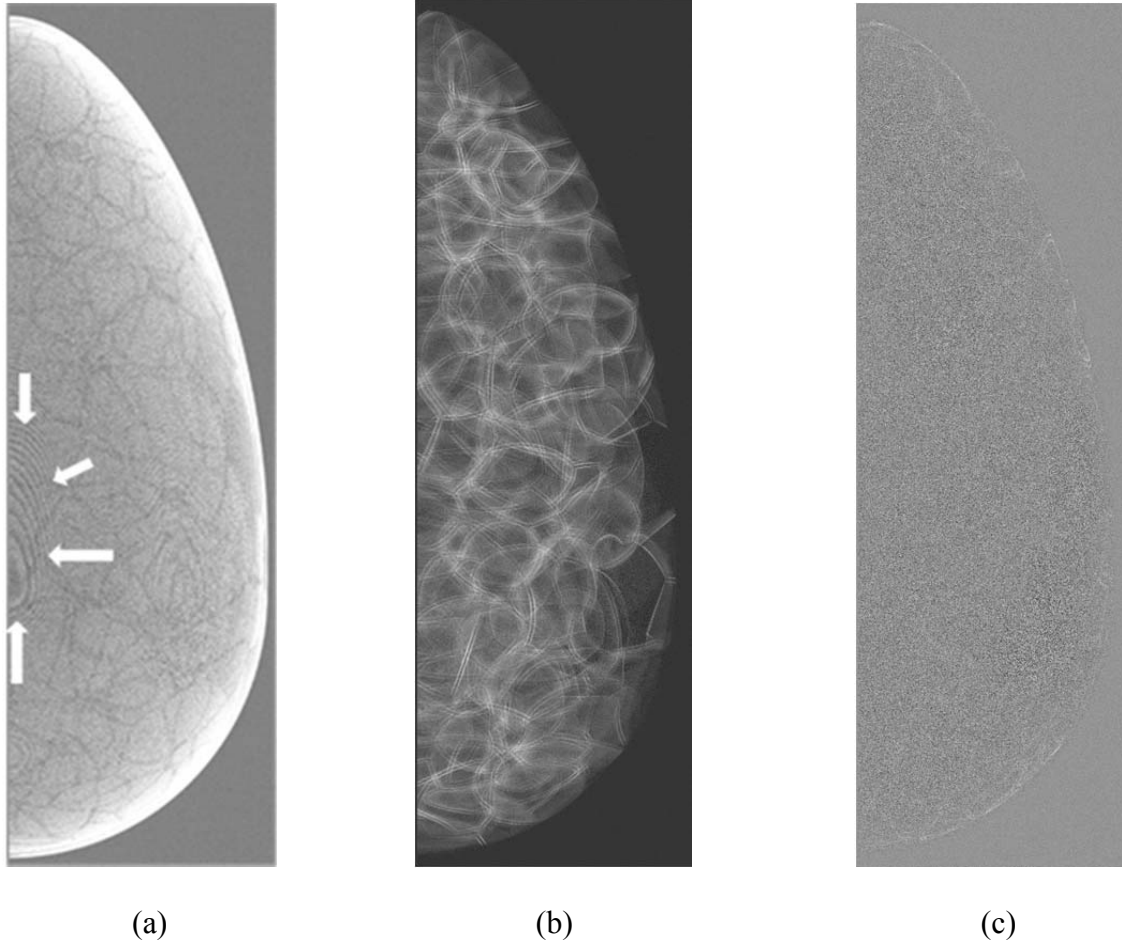
548   **Figure 10**: Illustration of the effect of PV in simulated projections of phantoms. Shown is the
549   contribution of (a) two-material voxels containing skin (cases 6-9; Table 2) (white arrows
550   indicate ripple artifacts); (b) two-material voxels containing ligaments (cases 10 and 11); and (c)
551   three-material voxels (cases 12 and 13).

552

553   **IV.B. Quantitative validation**

554   For quantitative validation of the proposed method, we utilized the Monte Carlo method, as
555   described in Section III.A. The number of points per voxel for the Monte Carlo simulation was
556   varied ($N_{MC} \in \{63, 500\}$). The number of repetitions was $N_{repeat}=100$. Table 4 contains the MSE_A
557   for PV of skin, PV of ligaments in two material voxels and PV of ligaments in three material
558   voxels, using the 200μm phantom shown in Fig. 8. MSE_A is computed using the Monte Carlo
559   method with $N_{MC}=63$ random points, using both methods discussed in section III.A.1.   The
560   corresponding estimated MSE_{MCR}, MSE_{MCS} are also shown.   Since we used q=6 bits for
561   representation of a partial volume percentage, the approximate quantization error, obtained using
562   Eq. (21), was MSE_q =2.09e-5.

563

564 Table 5 lists the numbers of two-material voxels containing skin (cases 6-9; see Table 2) and
565 ligaments (cases 10 and 11), three-material voxels (cases 12 and 13), as well as the average
566 execution times for one PV voxel using the linear approximation and the Monte Carlo method
567 with N_{MC}=63. Note that total execution time for the PV computation using the linear
568 approximation was about 37 times less than with the Monte Carlo method (N_{MC}=63).

569

570 **Table 4:** MSE_A, MSE_{MC} obtained using repetitions, and sample means for three types of voxels;
571 Monte Carlo method uses N_{MC}=63 points per PV voxel

572

Estimation method		Two material: Skin	Two material: ligaments	Three material: ligaments
MSE_{MCS}		1.797e-03	1.772e-03	1.574e-03
MSE_{MCR}		1.779e-03	1.754e-03	1.562e-03
MSE_A	Sample means	2.007e-05	4.323e-04	2.931e-04
	Repetition	3.797e-05	4.528e-04	3.058e-04

573

574 **Table 5**: The numbers of different types of voxels and the average execution times per PV voxel
575 for the proposed method and for the Monte Carlo simulation

576

		Voxel type	Two material: Skin	Two material: ligaments	Three materials
		Number of Voxels	1,597,042	6,435,881	87,610
Average execution time per PV voxel (ms)		Linear approximation	0.556	0.380	4.75
		Monte Carlo (N_{MC}=63)	0.371	21.6	22.1

577

578 Table 6 contains MSE_A for PV of skin, PV of ligaments in two material voxels and PV of
579 ligaments in three material voxels, on 200μm phantom shown in Fig. 8. MSE_A is computed using
580 the Monte Carlo method with N_{MC}=500 random points with the sample mean method. The
581 corresponding estimated MSE_{MCS} are also shown. Note that estimation of MSE_{MCR} was not
582 computationally feasible (the computation of MSE_{MCR} would require excessively large
583 computational time).

584

585 **Table 6:** MSE_A, MSE_{MC} obtained using the sample means method for three types of materials;
586 Monte Carlo method uses N_{MC}=500 points per PV voxel

587

Estimation method	Two material: Skin	Two material: ligaments	Three materials
MSE_{MCS}	2.260e-04	2.229e-04	1.984e-04
MSE_A	3.604e-05	4.487e-04	2.918e-04

588

589

590 To determine the influence of ligament boundary non-linearity on the accuracy of the proposed
591 PV estimation, we generated four 200µm phantoms corresponding to the classes discussed [49].
592 We varied the relative compartment orientation r_0, and the relative compartment size, r_s, as
593 shown in Table 7.

594

595 **Table 7:** Values of the input parameter defining relative compartment orientation, r_0, and the
596 relative compartment size, r_s, used for the generation of the four analyzed classes of phantoms.

597

r_s r_0	1	[0.01, 100]
1	*Class 1*	*Class 2*
[0.25, 4]	*Class 3*	*Class 4*

598

599 Table 8 contains MSE_A, MSE_{MC} obtained using the sample means method for three types of
600 materials on four phantoms with different non-linearity of ligament boundaries specified by
601 different ranges of r_s and r_0 (See Fig. 11); Monte Carlo method uses N_{MC}=63 points per PV
602 voxel.

603

604

605

606

607
608 (a)   (b)

609
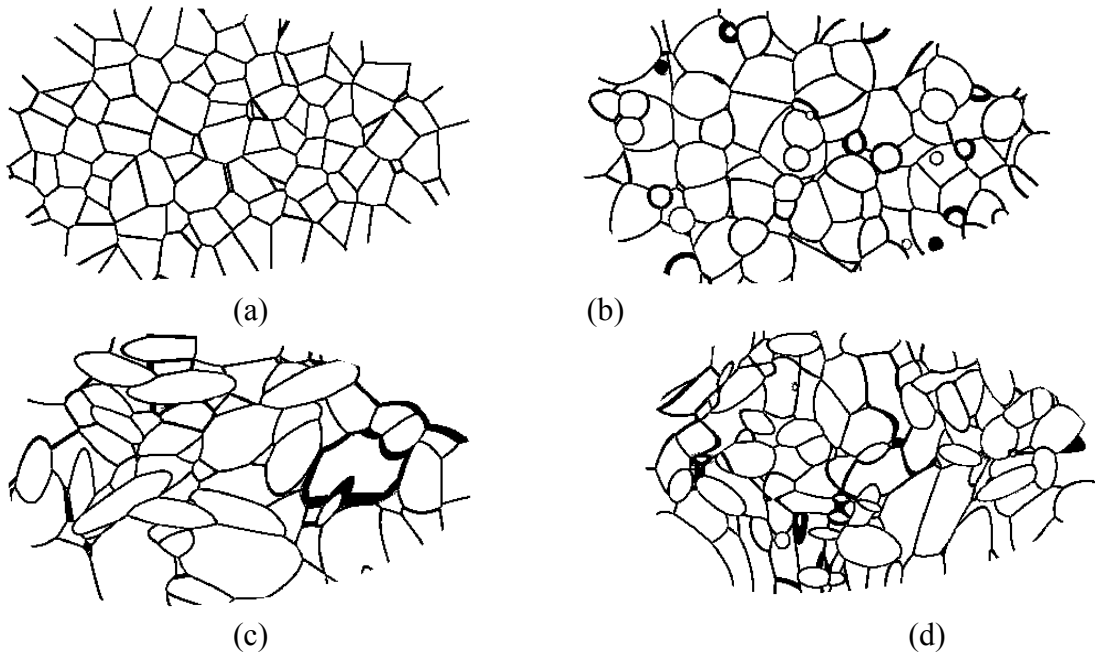610 (c)   (d)
611 **Figure 11:** Coronal cross-sections through sample phantoms from the four analyzed classes: (a)
612 Class 1 (b) Class 2; (3) Class 3; (d) Class 4 (see Table 7).

613

Table 8: *MSE_A, MSE_MC* for four phantoms with different non-linearity of ligament boundaries specified by different ranges of r_s and r_0. The Monte Carlo method uses N_{MC}=63 points per PV voxel.

Class 1	Estimation method	Two material: Skin	Two material: ligaments	Three materials
MSE_{MCS}		1.796e-03	1.764e-03	1.591e-03
MSE_A		1.983e-05	4.141e-05	2.571e-05

a) r_s=1, r_0=1

Class 2	Estimation method	Two material: Skin	Two material: ligaments	Three materials
MSE_{MCS}		1.796e-03	1.754e-03	1.573e-03
MSE_A		2.297e-05	3.698e-04	1.165e-04

b) r_s *in* [0.01, 100], r_0=1

Class 3	Estimation method	Two material: Skin	Two material: ligaments	Three materials
MSE_{MCS}		1.797e-03	1.814e-03	1.594e-03
MSE_A		1.528e-05	1.037e-03	3.305e-04

c) r_s =1, r_0 *in* [0.25, 4]

Class 4	Estimation method	Two material: Skin	Two material: ligaments	Three materials
MSE_{MCS}		1.796e-03	1.786e-03	1.578e-03
MSE_A		2.171e-05	2.144e-03	5.388e-04

d) r_s *in* [0.01, 100], r_0 *in* [0.25, 4]

V. Discussion

Fig. 8 suggests that the simulation of all 13 cases of PV voxels is qualitatively correct. The voxels containing two materials are detected at the boundaries of two materials (e.g., skin, compartment). The three material voxels are detected where the skin meets Cooper's ligaments and a compartment. Computed percentages (PVs) of skin and ligaments gradually decrease when departing from the inside of the skin (ligaments) (Fig. 8(b), 8(c)).

In Fig. 9(e), for a phantom with simulated PV, the equivalent x-ray attenuations of voxels on skin/air and ligaments/fat tissue boundaries were lower than the x-ray attenuation of dense tissue, hence the quantization artifacts were reduced and Cooper's ligaments and skin appeared thinner and their boundaries smoother in the projection (as compared to the phantom without PV, Fig. 9(d)). This is confirmed by the difference between projections of phantoms with two material PV voxels and with PV voxels containing skin (Fig. 10(b)).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

638   In Fig. 9(e), the characteristic stair-step quantization artifacts on tissue boundaries were
639   noticeably reduced with simulated PV. Simulation of two-material voxels with skin leads to
640   reduction of ripple artifacts due to sudden change of attenuation at the skin boundary. (Fig.
641   10(a)). Note that here we represent linear x-ray attenuation coefficient of a PV voxel as a
642   weighted average of the attenuation coefficients of materials contained in the voxel (instead of
643   using a single material attenuation coefficient). Hence, the proposed method can reduce aliasing
644   due to improved sampling of a continuous phantom. Comparison of Figs. 9(e) and 9(f) indicates
645   similar appearance of a phantom with PV simulated at a larger voxel size (200μm) to a phantom
646   simulated at a smaller voxel size(100μm) with no simulated PV. Note that a Cooper's ligament
647   in the lower central portion of Fig. 9(e) with simulated PV appears thinner, even when compared
648   with a smaller voxel size phantom without PV (Fig. 9(f)). Hence, the application of PV may lead
649   to an improvement in image quality without reducing voxel size. In comparison to noticeable
650   quality improvement of simulation of two-material PV voxels, the simulation of three-material
651   voxels leads to a relatively smaller improvement in image quality, by removing high-frequency
652   artifacts (see Fig. 10(c)).
653
654   The estimated accuracy of the PV computation (MSE_A) is better when using the proposed
655   approximation, than using the MC with 63 points, as indicated in Table 4. For skin, the accuracy
656   of approximation is close to the approximate quantization error MSE_q (calculated in the
657   beginning of Section IV.B.) The statistically insignificant discrepancy (2.09e-5 vs. 2.007e-5)
658   could be explained by error in estimating MSE_{MC}. Note that using N_{MC}=63 points per voxel in
659   Monte Carlo estimation corresponds to 6-bits resolution of the obtained PV estimates—the same
660   as the resolution due to discretization of the approximation.
661
662   Comparison of Table 4 and Table 6 shows that the estimate of MSE_A is stable (i.e., does not
663   change much) with increasing N_{MC}. Hence, the use in practice of relatively small N_{MC} (=63) to
664   estimate MSE_A is justified. When N_{MC} was increased, MSE_{MC} decreased (as expected from Eq.
665   (29)) and became comparable to MSE_A on two-material and three-material ligament voxels.
666   Observe, however, that this comparable accuracy is achieved at the expense of additional
667   computational time. Hence, our proposed approximation method is preferable for both fast and
668   accurate estimation of PV.
669
670   For a phantom with very linear ligament boundaries (Fig. 11(a)), MSE_A is very close to MSE_q
671   (see Table 8a). In this case, the linear approximation clearly has better accuracy than MC with
672   63 random points. As we can see, when r_0=1 but relative compartment sizes vary *in* [0.01, 100]
673   (Fig. 11(b)), the boundaries of ligaments are non-linear but relatively smooth. Hence, the MSE
674   of the linear approximation of two- or three-material ligaments are smaller than for the MSE of
675   the Monte Carlo (see Table 8b). Fig. 11(c) shows that when r_0 is allowed to vary in [0.25, 4] but
676   relative compartment size is r_s=1, the phantom's ligament boundaries are less linear. As a
677   consequence, MSE_A for ligament voxels are larger than in the previous case (Fig. 11(a) and
678   11(b)). Nevertheless, linear approximation still has smaller error than MC. Fig. 11(d) shows
679   that when r_0 vary in [0.25, 4] but relative compartment sizes r_s vary in [0.01, 100], ligament
680   boundaries are very non-linear. As a result, now MSE_A for ligaments is larger than MSE_{MC}. In
681   contrast, the volumes of three-material voxels are bounded by relatively smooth skin surface (in
682   addition to non-linear ligament surface), and in such cases linear approximation outperforms MC.

683 A better approximation (e.g., quadratic) may still be needed if the surfaces separating different
684 volumes are not linear, since the linear approximation on the ligament boundaries is not exact. If
685 the boundaries are highly non-linear, computation of the PV effect using the Monte Carlo (MC)
686 method may be a better choice (since MC provides a controllable approximation error) on fast
687 enough hardware/implementation.

688

689 Our algorithms for two and three materials are very efficient. In the two-material case,
690 Algorithm A1 can be used to compute partial volume by solving one inner product and the
691 volume of a few geometric primitives. In the three-material case, Algorithm A3 has converted
692 the 3D volume problem into a 2D area problem using the Gauss-Ostrogradsky theorem.
693 Moreover, Algorithm A3 can be used for any number of materials, but for two-material cases,
694 Algorithm A1 is faster.

695

696 The obtained average execution times of PV estimation per voxel, Table 5, are platform and
697 implementation dependent. Using the Monte Carlo method for two-material voxels containing
698 skin (cases 6-9) is relatively straightforward resulting in smaller average execution times than
699 using our implementation of the proposed algorithm. Note that while being conceptually simple,
700 estimation using the Monte Carlo method of voxels containing ligament tissue relies upon
701 computation of distance from the median surface which is resource intensive (it reduces to
702 numerical solution of the polynomial equation of 6-th degree [44]). Since PV with two-materials
703 containing ligaments (cases 10-11) are predominant in the considered phantom, the total time to
704 compute partial volumes using the linear approximation was much smaller than using the Monte
705 Carlo method with N_{MC}=63 points. Therefore, using the linear approximation should be faster
706 than the Monte Carlo method. On the other hand, unlike the linear approximation, where the
707 accuracy depends on the linearity of the material boundaries, the accuracy of the Monte Carlo
708 method can be controlled (by choosing large enough N_{MC}, see Eq. (29)). Hence, if the accurate
709 computation of partial volumes separated by highly non-linear surfaces is necessary, or if the
710 application/platform (e.g., parallel platforms) where the Monte Carlo method is efficient are
711 available, the Monte Carlo may be a method of choice for computing partial volumes. The
712 determination of smallest N_{MC}, for a given MSE_{MC} is part of our work in progress.

713

714 Note that for realistic cases of non-linear ligament boundaries the estimated MSE_A of two-
715 material voxels containing ligaments and of three-material voxels were larger than the
716 quantization error MSE_q when q=6 bits are used. Hence, it is sufficient to use 6 bits to represent
717 partial volumes computed using the proposed linear approximation. Note that in [38] we
718 proposed using 7 bits per partial volume.

719

720 Observe that the proposed scheme reserves 4 bits to encode material type; currently we utilize 2
721 bits. In this way, we have the potential to include more material types and their combinations
722 (e.g., lesions, calcifications, ducts, etc).

723

724 Note, finally, that we have proposed three techniques to estimate MSE_{MC}, and experimentally
725 confirmed their similar behavior in the proposed application. The third technique is based upon
726 the estimation of sample means of computed PV_{MC}, which avoids the estimation of ε_{MC}. This
727 technique does not require knowledge of true values of PV_j at each voxel nor the availability of
728 distribution $p(PV)$, and could be thus potentially used in other estimation problems.

VI. Conclusions

We have developed a method for simulating PV in software breast phantom voxels, which contains multiple simulated tissues. The percentage of simulated tissues was estimated using a planar approximation of the boundary between different tissue regions, based upon the segmentation into geometric primitives and the Gauss-Ostrogradsky theorem. A quantitative assessment of the planar approximation using Monte Carlo estimation of computed PV showed satisfactory accuracy of the proposed method. A qualitative comparison of simulated mammographic projections confirmed that the PV simulation can improve the image quality by reducing the quantization artifacts. A future work would involve human or model observer studies to quantify the image quality improvement.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

749 **Appendix 1: Pseudocode of algorithms for computing partial volume of a voxel above a**
750 **plane and above two planes**
751
752 Vi=PV_2($\Delta x, \mathbf{P}_i, i=1,...,8, \mathbf{N}, \mathbf{X}_0$)
753 // **Inputs:** voxel linear dimension Δx; voxel vertices $\mathbf{P}_i, i=1,...,8$
754 // a plane normal **N**; a point $\mathbf{x}_0$ on the plane;
755 // **Output:** Partial volume Vi of the voxel above the plane
756 COMPUTE *nVertex*
757 IF *nVertex*>4
758   DETERMINE voxel vertices $\mathbf{Q}_i, i=1,...,8-nVertex$ satisfying $(\mathbf{Q}_i-\mathbf{x}_0)\cdot\mathbf{N}\leq 0$
759   RETURN $\Delta x^3 - PV\_compute\_2$($\mathbf{Q}_i, i=1,...8-nVertex, \Delta x, -\mathbf{N}, \mathbf{x}_0, 8-nVertex$)
760 ELSE
761   DETERMINE voxel vertices $\mathbf{Q}_i, i=1,...,nVertex$ satisfying $(\mathbf{Q}_i-\mathbf{x}_0)\cdot\mathbf{N}>0$
762   RETURN $PV\_compute\_2$($\mathbf{Q}_i, i=1,...nVertex, \Delta x, \mathbf{N}, \mathbf{x}_0, nVertex$);
763
764 **Algorithm A1**: Conceptual algorithm for computing partial volume of a voxel above a given
765 plane
766

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```
767    Vi= PV_compute_2 ( Qi,i=1,...nVertex,Δx,N, x0,nVertex )
768    // Inputs: Voxel vertices Qi,i=1,…,nVertex above plane π (nVertex≤4!)
769    // voxel linear dimension Δx
770    // normal N and a point x0 specifying plane π.
771    // Output: Partial volume Vi of the voxel above the plane π
772    IF nVertex==0
773           RETURN 0.    //Partial volume is 0
774    ELSEIF nVertex==1    //Volume is a right angle triangular pyramid. CASE A (see Table 3)
775              COMPUTE nPoint=2*nVertex+1. //the number of intersections.
776              COMPUTE intersections P, Q, R between π and ei. the edges containing Q1.
777              COMPUTE the distances between intersections and Q1.
778              RETURN volume of a tetrahedron PQRQ1. //(See Fig.4(a))
779    ELSEIF nVertex==2
780               DETERMINE the edge e containing Q1 and Q2
781               IF N⊥e        //Volume Vi is a triangular prism. CASE B1
782                  COMPUTE intersections P, Q between π and edges containing Q1 (except
783                  e).
784                  RETURN volume of a prism defined with base PQQ1 and height e.
785                  //(See Fig. 4(b))
786               ELSE   // volume is a triangular cut pyramid CASE B2
787                  COMPUTE nPoint=2*nVertex+1.  //the number of intersections.
788                  COMPUTE intersections P, Q, R, S, T between π and each edge (or
789                  extension of edge) containing Q1 or Q2.
790                  RETURN volume difference b/w tetrahedra PQRQ1, PSTQ2.//See Fig. 4(c)
791               ENDIF
792    ELSEIF  nVertex==3  //"double cut" pyramid. CASE C.
793              COMPUTE nPoint=2*nVertex+1. // the number of intersections.
794              COMPUTE intersections P, Q, R, S, T, U, W between π and each edge (or
795              extension of edge) containing Q1, Q2 or Q3.
796              RETURN volume difference b/w tetrahedra PQRQ1, PSTQ2, QUWQ3 .//See Fig.4(d)
797    ELSEIF  nVertex==4
798               IF vertices Qi i=1,…,4 are coplanar //The volume Vi is a prismoid.
799                    //CASE D1
800                  COMPUTE intersections P,Q,R,S between π and edges vertical to plane
801                  defined by Qi i=1,…,4.
802                  RETURN volume of prismoid Q1Q2Q3Q4PQRS.              //See Fig.4(e)
803               ELSE   //"triple cut" pyramid. CASE D2
804                  COMPUTE nPoint=2*nVertex+1.   //The number of intersections.
805                  COMPUTE intersections P, Q, R, S, T, U, W, Y, Z between π and each
806                  edge (or extension of edge) containing Q1, Q2, Q3 or Q4.
807                  RETURN volume difference b/w tetrahedron PQRQ2 and tetrahedra PSTQ1,
808                  QUWQ3, and RYZQ4.       //See Fig. 4(f))
809               ENDIF
810    ENDIF
811
```

Algorithm A2: Algorithm for computation of partial volume of a voxel above a plane for different number of vertices above the plane

815   Vi=*PV_3*(Δx,$\mathbf{P}_i, i = 1,...,8, \mathbf{x}_1, \mathbf{N}_1, \mathbf{x}_2, \mathbf{N}_2$)

816   // **Inputs:** voxel linear dimension Δx;

817   // voxel vertices $\mathbf{P}_i$, i=1,…,8

818   // plane normals $\mathbf{N}_1, \mathbf{N}_2$ of planes $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$; points $\mathbf{x}_1$, $\mathbf{x}_2$ on the planes

819   // **Output:** Partial volume *Vi* of the voxel above the planes $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$

820

821   COMPUTE *nVertex* as the number of vertices $\mathbf{P}_i$ satisfying $(\mathbf{P}_i - \mathbf{x}_1) \cdot \mathbf{N}_1 > 0$ and $(\mathbf{P}_i - \mathbf{x}_2) \cdot \mathbf{N}_2 > 0$.

822   IF   *nVertex*==8

823   RETURN Δx^3

824   ELSEIF *nVertex*~=0

825   RETURN PV_compute_3(Δx,$\mathbf{x}_1, \mathbf{N}_1, \mathbf{x}_2, \mathbf{N}_2, \mathbf{P}_i, i = 1,...,8$).

826   ELSE

827   COMPUTE *nVertex1* and *nVertex2*, the number of vertices above $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$.

828   IF *nVertex1* * *nVertex2*~=0

829   // there is no vertex above both planes, but there is a vertex above each plane.

830   RETURN *PV_2*(Δx,$\mathbf{P}_i, i = 1,...,8, \mathbf{N}_1, \mathbf{x}_1$) -*PV_compute_3*($\Delta$x,$\mathbf{x}_1, \mathbf{N}_1, \mathbf{x}_2, -\mathbf{N}_2, \mathbf{P}_i, i = 1,...,8$). //(See Fig.

831   7)

832   ELSE

833   RETURN 0.

834   ENDIF

835   ENDIF

836

837

838   Vi=*PV_compute_3*(Δx,$\mathbf{x}_1, \mathbf{N}_1, \mathbf{x}_2, \mathbf{N}_2, \mathbf{P}_i, i = 1,...,8$)

839   // **Inputs:** voxel linear dimension Δx;

840   // plane normals $\mathbf{N}_1, \mathbf{N}_2$ of planes $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$; points $\mathbf{x}_1$, $\mathbf{x}_2$ on the planes

841   // Voxel vertices $\mathbf{P}_i, i = 1,…,8$ such that at least one vertex is above both planes $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$

842   // **Output:** Partial volume *Vi* of the voxel above the planes $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$

843   CHOOSE any vertex of the voxel above the two planes, denoted by $\mathbf{Q}_1$.

844   FOR each voxel side $\boldsymbol{\sigma}_i$, i=1,…,8 that does not contain $\mathbf{Q}_1$

845   DETERMINE the set S_{σ_i} of points above or on $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ belonging to one of the

846   following sets: intersections between edges of $\boldsymbol{\sigma}_i$ and planes $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$;

847   intersections between $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$ and $\boldsymbol{\sigma}_i$; vertices belonging to $\boldsymbol{\sigma}_i$.

848   COMPUTE S_i as the area of convex polygon with vertices from S_{σ_i}.

849   COMPUTE the distances d_j between $\mathbf{Q}_1$ and $\boldsymbol{\pi}_j, j = 1,2$.

850   FOR each plane $\boldsymbol{\pi}_j, j = 1,2$

851   COMPUTE a set S_{π_j} belonging to one of the following sets: intersections between

852          $\boldsymbol{\pi}_j$ and the voxel edges that are above or on $\boldsymbol{\pi}_k, k \neq j$; intersections of $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$ and

853   the surface of the voxel.

854   COMPUTE A_j as the area of a convex polygon with vertices from S_{π_j}.

855   RETURN Vi=$\frac{1}{3}[(S_1 + S_2 + S_3)\Delta x + A_1 d_1 + A_2 d_2]$;

856

857   **Algorithm A3:** Algorithm for computation of partial volume of a voxel above two planes

858

859

860   **Appendix 2: Some properties of Monte Carlo estimation of partial volume**

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

861 Observe that the true value of a partial volume PV is the probability that a randomly chosen
862 point during the Monte Carlo computation of partial volume is within the volume of interest.
863 Hence, the count of points (N_{MCI} in Eq. (16)) follows a Binomial distribution with expectation
864 $N_{MC} \cdot PV$ and variance $N_{MC} \cdot PV \cdot (1 - PV)$ [43]. From this and Eq. (16), (17), follows:

865 $\quad E(\varepsilon_{MC}|PV)=0$ $\hspace{6cm}$ (A1)

866 $\quad E(\varepsilon^2_{MC}|PV)=\dfrac{PV\,(1-PV)}{N_{MC}}$ $\hspace{5cm}$ (A2)

867 Note that it is suitable to treat the true value of the partial volume, PV, as a random variable
868 (since it varies throughout the phantom in fashion unknown to the algorithm for PV estimation).
869 Using the conditional expectation $E(\varepsilon^2_{MC}|PV)$, the expectation $E(\varepsilon^2_{MC})$of the error of the
870 Monte Carlo method can be expressed as [43]:

871 $\quad E(\varepsilon_{MC}^2) = \int_0^1 E(\varepsilon_{MC}^2|PV)p(PV)dPV$ $\hspace{4cm}$ (A3)

872 By combining Eq. (A2) and (A3) we can easily obtain:

873 $\quad E(\varepsilon_{MC}^2) = \int_0^1 \dfrac{PV(1-PV)}{N_{MC}}p(PV)dPV = \dfrac{1}{N_{MC}}\left(E(PV) - E(PV^2)\right)$ $\hspace{1cm}$ (A4)

874 Also, similarly, using Eq. (A1) we can obtain:

875 $\quad E(\varepsilon_{MC} * PV) = \int_0^1 E(\varepsilon_{MC}|PV) \cdot PV \cdot p(PV)dPV = 0$ $\hspace{2cm}$ (A5)

876 $\quad E(\varepsilon_{MC}) = \int_0^1 E(\varepsilon_{MC}|PV) \cdot p(PV)dPV = 0$ $\hspace{3cm}$ (A6)

877 Due to Eq. (A6),

878 $\quad E(PV)=E(PV_{MC})$. $\hspace{5cm}$ (A7)

879 Note that PV_A is a function of PV (in the ideal case, $PV_A=PV$!) and therefore $\varepsilon_A= \varepsilon_A(PV)$.
880 Due to Eq. (A1):

881 $\quad E(\varepsilon_{MC}\varepsilon_A) = \int E(\varepsilon_{MC}|PV)\,\varepsilon_A(PV)dPV = 0$ $\hspace{3cm}$ (A8)

882 i.e., the approximation error and the error of Monte Carlo method are not correlated. Note that
883 Eq. (A8) holds for linear and non-linear approximation methods.

884

885

886 **References**

887 [1]   P. Taylor and R. Owens, "Simulated Mammography using Synthetic 3D Breasts," in 4th
888   International Workshop on Digital Mammography, Nejmegen, Netherlands, 1998, pp. 283-290.
889 [2]   P. R. Bakic, et al., "Mammogram synthesis using a 3D simulation. I. Breast tissue model and
890   image acquisition simulation," Medical Physics, vol. 29, pp. 2131-9, 2002.

891 [3]   P. R. Bakic, et al., "Mammogram synthesis using a 3D simulation. II. Evaluation of synthetic
892   mammogram texture," Medical Physics, vol. 29, pp. 2140-51, 2002.
893 [4]   P. R. Bakic, et al., "Mammogram synthesis using a three-dimensional simulation. III. Modeling
894   and evaluation of the breast ductal network," Medical Physics, vol. 30, pp. 1914-25, 2003.
895 [5]   K. Bliznakova, et al., "A three-dimensional breast software phantom for mammography
896   simulation," Physics in Medicine & Biology, vol. 48, pp. 3699-719, 2003.
897 [6]   C. Zhang, et al., "Development of an Anthropomorphic Breast Software Phantom Based on
898   Region Growing Algorithm," in SPIE Medical Imaging, San Diego, CA, 2008.
899 [7]   K. Bliznakova, et al., "Evaluation of an improved algorithm for producing realistic 3D breast
900   software phantoms: Application for mammography," Medical Physics, vol. 37, pp. 5604-5617,
901   2010.
902 [8]   I. Reiser, et al., "Toward validation of a 3D structured background model for breast imaging,"
903   presented at the Physics of Medical Imaging, San Diego, CA, 2010.
904 [9]   B. Chen, et al., "An anthropomorphic breast model for breast imaging simulation and
905   optimization," Academic Radiology, vol. 18, pp. 536-546, 2011.
906 [10] A. B. Lau, et al., "A statistically defined anthropomorphic software breast phantom " Medical
907   Physics, vol. 39, pp. 3375-3385, 2012.
908 [11] D. D. Pokrajac, et al., "Optimized generation of high resolution breast anthropomorphic software
909   phantoms," Medical Physics, vol. 39, pp. 2290-2302, 2012.
910 [12] C. Hoeschen, et al., "A high resolution voxel phantom of the breast for dose calculations in
911   mammography," Radiation Protection Dosimetry vol. 114, pp. 406-409, 2005.
912 [13] J. M. O'Connor, et al., "Comparison of two methods to develop breast models for simulation of
913   breast tomosynthesis and CT," in Digital Mammography (IWDM). vol. 5116, E. A. Krupinski, Ed.,
914   ed Berlin-Heidelberg: Springer-Verlag, 2008, pp. 417-425.
915 [14] C. M. Li, et al., "Methodology for generating a 3D computerized breast phantom from empirical
916   data," Medical Physics, vol. 36, pp. 3122-31, 2009.
917 [15] J. M. O'Connor, et al., "Development of an Ensemble of Digital Breast Object Models," in Digital
918   Mammography (IWDM). vol. 6136, J. Marti, Ed., ed Berlin-Heidelberg: Springer-Verlag, 2010, pp.
919   54-61.
920 [16] N. Kiarashi, et al., "Development of matched virtual and physical breast phantoms based on
921   patient data," presented at the Physics of Medical Imaging, Lake Buena Vista, FL, 2013.
922 [17] J. M. O'Connor, et al., "Generation of voxelized breast phantoms from surgical mastectomy
923   specimens " Medical Physics, vol. 40, p. 041915, 2013.
924 [18] P. R. Bakic, et al., "Validation and Optimization of Digital Breast Tomosynthesis Reconstruction
925   using an Anthropomorphic Software Breast Phantom," presented at the SPIE Medical Imaging:
926   Physics of Medical Imaging San Diego, CA, 2010.
927 [19] P. R. Bakic, et al., "Analysis of Geometric Accuracy in Digital Breast Tomosynthesis
928   Reconstruction," presented at the Int'l Workshop on Digital Mammography, Girona, Spain, 2010.
929 [20] R. Zeng, et al., "Is the outcome of optimizing the system acquisition parameters sensitive to the
930   reconstruction algorithm in Digital Breast Tomosynthesis?," presented at the International
931   Workshop on Breast Imaging (IWDM), Philadelphia, PA, 2012.
932 [21] M. A. C. Vieira, et al., "Effect of denoising on the quality of reconstructed images in digital breast
933   tomosynthesis," presented at the Physics of Medical Imaging, Lake Buena Vista, FL, 2013.
934 [22] M. A. C. Vieira, et al., "Filtering of Poisson Noise in Digital Mammography Using Local Statistics
935   and Adaptive Wiener Filter," presented at the International Workshop on Breast Imaging
936   (IWDM), Philadelphia, PA, 2012.
937 [23] P. R. Bakic, et al., "Comparison of 3D and 2D Breast Density Estimation from Synthetic
938   Ultrasound Tomography Images and Digital Mammograms of Anthropomorphic Software Breast

939 Phantoms," presented at the SPIE Medical Imaging: Physics of Medical Imaging Lake Buena
940 Vista, FL, 2011.

941 [24] H. Yang, et al., "Performance analysis of EM-MPM and K-means clustering in 3D ultrasound
942 image segmentation " presented at the 2012 IEEE International Conference on
943 Electro/Information Technology, EIT 2012, Indianapolis, IN, 2012.

944 [25] A. B. Lau, et al., " An Anthropomorphic Software Breast Phantom for Tomosynthesis Simulation:
945 Power Spectrum Analysis of Phantom Reconstructions," Medical Physics, vol. 37, p. 3473, 2010.

946 [26] P. R. Bakic, et al., "An Anthropomorphic Software Breast Phantom for Tomosynthesis Simulation:
947 Power Spectrum Analysis of Phantom Projections," presented at the Int'l Workshop on Digital
948 Mammography, Girona, Spain, 2010.

949 [27] D. Kontos, et al., "Evaluating the Effect of Tomosynthesis Acquisition Parameters on Image
950 Texture: A Study Based on an Anthropomorphic Breast Tissue Software Model," in 9th
951 International Workshop on Digital Mammography, Tucson, AZ, 2008.

952 [28] P. R. Bakic, et al., "Testing Realism of Software Breast Phantoms: Texture Analysis of Synthetic
953 Mammograms " presented at the Physics of Medical Imaging, Lake Buena Vista, FL, 2013.

954 [29] S. Young, et al., "Estimating DBT performance in detection tasks with variable-background
955 phantoms," presented at the SPIE Medical Imaging: Physics of Medical imaging, Lake Buena
956 Vista, FL, 2009.

957 [30] A. B. Lau, et al., "Towards Visual-Search Model Observers for Mass Detection in Breast
958 Tomosynthesis," presented at the Physics of Medical Imaging, Lake Buena Vista, FL, 2013.

959 [31] S. Young, et al., "A virtual trial framework for quantifying the detectability of masses in breast
960 tomosynthesis projection data," Medical Physics, vol. 40, p. 051914, 2013.

961 [32] A.-K. Carton, et al., "Development of a 3D physical anthropomorphic breast phantom," in
962 Tomosynthesis Imaging Symposium 2009: Frontiers in Research and Clinical Applications,
963 Durham, NC, 2009.

964 [33] A.-K. Carton, et al., "Development of a 3D high-resolution physical anthropomorphic breast
965 phantom," presented at the SPIE Medical Imaging: Physics of Medical Imaging, San Diego, CA,
966 2010.

967 [34] A.-K. Carton, et al., "Development of a physical 3D anthropomorphic breast phantom.," Medical
968 Physics, vol. 38, pp. 891-896, 2011.

969 [35] C. C. Brunner, et al., "Evaluation of various mammography phantoms for image quality
970 assessment in Digital Breast Tomosynthesis," presented at the International Workshop on Breast
971 Imaging (IWDM), Philadephia, PA, 2012.

972 [36] R. Karunamuni, et al., "Exploring the relationship between SDNR and detectability in dual-energy
973 breast x-ray imaging," presented at the Physics of Medical Imaging, Lake Buena Vista, FL, 2013.

974 [37] F. Chen, et al., "Partial Volume Simulation in Software Breast Phantoms," presented at the
975 Physics of Medical Imaging, San Diego, CA, 2012.

976 [38] F. Chen, et al., "Simulation of Three Material Partial Volume Averaging in a Software Breast
977 Phantom," presented at the International Workshop on Breast Imaging (IWDM), Philadelphia, PA,
978 2012.

979 [39] P. R. Bakic, et al., "Development and Characterization of an Anthropomorphic Breast Software
980 Phantom Based upon Region-Growing Algorithm," Medical Physics, vol. 38, pp. 3165-3176, 2011.

981 [40] C. M. Li, et al., "Computerized 3D Breast Phantom with Enhanced High-Resolution Detail,"
982 presented at the SPIE Medical Imaging: Physics of Medical Imaging Lake Buena Vista, FL 2009.

983 [41] G. B. Folland, Advanced Calculus. Upper Saddle River, NJ: Prentice-Hall, Inc., 2002.

984 [42] G. Thomas, Calculus & Analytic Geometry: Dorling Kindersley Pvt Ltd., 2010.

985 [43] R. Y. Rubinstein, Kroese, D. P., Simulation and the Monte Carlo Method, 2007.

986 [44] *P. J. Schneider and D. H. Eberly. (2008, 11/21/2013). Distance from point to a general quadratic*
987 *curve   or   a   general   quadratic   surface.   Available:*
988 *http://www.geometrictools.com/Documentation/DistancePointToQuadratic.pdf*
989 [45] *K. Pohlmann, Principles of Digital Audio: McGraw-Hill/TAB Electronics, 2010.*
990 [46] *A. Ringberg, et al., "Of cup and bra size: Reply to a prospective study of breast size and*
991 *premenopausal breast cancer incidence," International Journal of Cancer, vol. 119, pp. 2242-*
992 *2243, 2006.*
993 [47] *T. L. Pope, et al., "Breast skin thickness: Normal range and cause of thickening shown on film-*
994 *screen mammography," J. Canadian Assoc Radiol, vol. 35, pp. 365-368, 1984.*
995 [48] *H. Ulger, et al., "Effect of age, breast size, menopausal and hormonal status on mammographic*
996 *skin thickness," Skin Res. Technol., vol. 9, pp. 284-289, 2003.*
997 [49] *J. M. L. F. Contijoch, D. D. Pokrajac, A. D. A. Maidment, P. R. Bakic "Shape analysis of simulated*
998 *breast anatomical structures," 2012.*
999 [50] *N. V. Ruiter, et al., "Model-based registration of x-ray mammograms and MR images of the*
1000 *female breast," IEEE Trans. Nuclear Science, vol. 53, pp. 204-211, 2006.*
1001 [51] *J. H. Hubbel and S. M. Seltzer. (1996, 11/21/2013). Tables of X-Ray Mass Attenuation Coefficients*
1002 *and   Mass   Energy-Absorption   Coefficients   (version   1.4).   Available:*
1003 *http://www.nist.gov/pml/data/xraycoef/index.cfm*

1004

1005